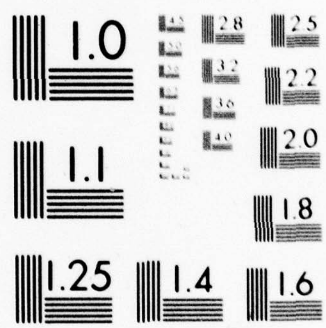
AD-A072 259 CALIFORNIA UNIV LOS ANGELES SCHOOL OF ENGINEERING A--ETC F/G 1/2
ENTROPHY ANALYSIS OF FEEDBACK FLIGHT DYNAMIC CONTROL SYSTEMS.(U)
JAN 79 H L WEIDEMANN, C T LEONDES F33615-77-C-3013
UNCLASSIFIED AFFDL-TR-78-123 NL

1 OF 2
AD A072259



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

AFFDL-TR-78-123

LEVEL

AD A072259

ENTROPY ANALYSIS OF FEEDBACK FLIGHT DYNAMIC CONTROL SYSTEMS

UNIVERSITY OF CALIFORNIA, LOS ANGELES
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
7620 BOELTER HALL, UCLA
LOS ANGELES, CALIFORNIA 90024

January 1979

TECHNICAL REPORT AFFDL-TR-78-123
Final Report for Period May 1977-May 1978

Approved for public release; distribution unlimited.

AIR FORCE FLIGHT DYNAMICS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

79 08 1 016

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

Gerald W. Francis
Gerald W. Francis

James W. Morris
JAMES W. MORRIS, Chief
Control Systems Development Branch
Flight Control Division

FOR THE COMMANDER

M. A. Ostgaard
MORRIS A. OSTGAARD
Assistant for Research and
Technology
Flight Control Division

"If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify AFFDL/FGL ,W-PAFB, OH 45433 to help us maintain a current mailing list".

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

AIR FORCE/56780/6 July 1979 — 100

(18) AFFDL (19) TR-78-123

4. TITLE (and Subtitle) (6) ENTROPHY ANALYSIS OF FEEDBACK FLIGHT DYNAMIC CONTROL SYSTEMS,	5. TYPE OF REPORT & PERIOD COVERED FINAL REPORT 5/77-5/78
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Henry Lawrence Weidemann ~~(Principal Investigator,~~ C. T. Leondes)	8. CONTRACT OR GRANT NUMBER(s) (15) F33615-77-C-3013
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of California, Los Angeles School of Engineering and Applied Science 7620 Boelter Hall, UCLA, Los Angeles, CA 90024	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 2307-03-02 (17) 03
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Systems Command ASD/ FGL WPAFB, OH 45433 Gerald W. Francis Gerald W. Francis	12. REPORT DATE 5/1/78
	13. NUMBER OF PAGES 154
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research 1030 E. Green Street Pasadena, CA 91106 Attn: Mr. Perry Beilke	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

(11) Jan 79

Approved for public release, distribution unlimited

(12) 165 p.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(9) Final rept. May 77-May 78,

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Entropy function, Flight Control Systems, Aircraft Sensors

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report is a study of the application of the entropy function of Information Theory to the analysis of sampled data systems so characteristic of the evolving and important field of digital flight control, including multimode systems. The systems studied are both feed-forward and feedback with the emphasis placed on the regulator problem. The feature common to all the configurations is the presence of a sensor which measures the input signal and which has an output that is usually some random function of the input. For the purposes of the analysis it is convenient to describe the behavior of the sensor by its next pg

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

404 637 79 08 1 016

Sensor Channel Transmittance, which is defined as the mutual information between the input and output of the measuring device. This quantity is not independent of the properties of the input signal; however, in any given problem it need only be calculated once.

For proofs relating to the estimation problem the constraints on the system elements are so relaxed that the sensor model need not be known and the only description of this device that is required is its Sensor Channel Transmittance. This implies that processors not having acceptable models, such as human operators, may now be successfully studied through the use of entropy analysis. For the feedback problem the restrictions are somewhat tighter in that although the sensor data processing may be nonlinear the noise must be constrained to be additive. However, this class of sensors is very important and the theory of feedback control is advanced significantly.

The major result of the report is that when a sensor path (either feedforward or feedback) is used to improve the performance of a system such as a flight control system, the entropy of the system error can never be reduced by more than an amount equal to the Sensor Channel Transmittance. This approach leads to the determination of the optimum system performance by using only open loop quantities that are easily determined. None of the results involve calculating the optimum filter that must be used to achieve the minimum entropy performance.

This research is important because it imbeds the control problem in the communication problem and clearly demonstrates the manner in which the information handling capability of the system elements limits performance, and is, therefore, of considerable potential significance to advanced flight dynamic systems.

PREFACE

This report is a study of the application of the entropy function of Information Theory to the analysis of sampled data systems so characteristic of the evolving and important field of digital control, including multimode systems. The systems studied are both feed-forward and feeback with the emphasis placed on the regulator problem. The feature common to all the configurations is the presence of a sensor which measures the input signal and which has an output that is usually some random function of the input. For the purposes of the analysis it is convenient to describe the behavior of the sensor by its Sensor Channel Transmittance, which is defined as the mutual information between the input and output of the measuring device. This quantity is not independent of the properties of the input signal; however, in any given problem it need only be calculated once.

For proofs relating to the estimation problem the constraints on the system elements are so relaxed that the sensor model need not be known and the only description of this device that is required is its Sensor Channel Transmittance. This implies that processors not having acceptable models, such as human operators, may now be successfully studied through the use of entropy analysis. For the feedback problem the restrictions are somewhattighter in that although the sensor data

processing may be nonlinear the noise must be constrained to be additive. However this class of sensors is very important and the theory of feedback control is advanced significantly.

The major result of the report is that when a sensor path (either feed-forward of feedback) is used to improve the performance of a system such as a flight control system, the entropy of the system error can never be reduced by more than an amount equal to the Sensor Channel Transmittance. This approach leads to the determinationof the optimum system performance by using only open loop quantities that are easily determined. None of the results involve claculating the optimum filter that must be used to achieve the minimum entropy performance.

This research is important because it imbeds the control problem in the communication problem and clearly deomstrates the manner in which the information handling capability of the system elements limits performance, and is therefore, of considerable potential significance to the advanced flight dynamic systems.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued).

LIST OF ILLUSTRATIONS

CHAPTER ONE
INTRODUCTION

1.1 Historical Survey

Historically, the theory of information began with Claude Shannon's publication of "The Mathematical Theory of Communications" [1]. This fundamental work is so profound that even though many authors are able to prove the Shannon theorems more simply [2,3], or supply proofs where none were given [4,5,6], or merely reinterpret the theory in a more useful manner [7,8,9], it is very rare to find published results that are not based in some way on a remark or idea generated by Shannon. Because the theory is so young, much of the significant published research has been directed toward developing and understanding the mathematical techniques [10,11,12, 13,14,15,16,17] rather than extending the theory to other applications. However, this survey, and in fact this entire dissertation, will be primarily interested in the applications of this theory which by now appears to have a very rigorous foundation.

The principles of present day information theory are the results of attempts to solve the very basic problem of calculating the amount of information contained in various random, but not very simple, objects. This measure of the information content, called entropy, provides, at the very least, a criteron for evaluating any encoding procedure [18]. In fact, the usual units of entropy, bits, indicates the preoccupation of the present theory with symbols and codes.

For the most part, information theory has remained the private preserve of the coding theorists, and it is perhaps unfortunate that the effect of the need for high powered space-age coding analysis has been to diminish the true universal value of entropy as a measure of information. It is rather apparent that there has been relatively little conclusive work applying the concepts of entropy to other problem areas. Nevertheless, a few valuable applications, generalizations, and interpretations of the theory have been obtained, with the most useful results appearing for:

a. <u>The Design of Experiments</u> [19,20]. DeGroot's effort [20] in this field is even more remarkable than is first apparent because useful results regarding the sequential design of experiments may be obtained from a very general axiomatic definition of uncertainty. This implies that for other areas of interest (filtering, controlling, etc.) the specific measures of uncertainty such as entropy and variance may be interchangeable.

b. <u>One Armed Bandit Problems</u> [21]. Kelly's approach to a gambling problem [21] is probably one of the earliest non-coding applications of entropy. Since gambling, economics, filtering, etc., all have a common probabilistic basis it will not be very surprising if derivations such as those made by Kelly, will be made in other fields.

c. <u>Human Operator Systems</u> [22]. Elkind's paper [22] is an attempt to apply entropy concepts to a type of system (human operator) that is not easily analyzed using existing techniques. This work must definitely be considered a precursor of the entropy analysis of

automatic control systems. However because his analysis depends on the information transmitted by the closed loop system, Elkind overlooks two important facets of the problem:

i) The information transmitted by the operator as an individual component of the whole system is not known.

ii) The effect of the limited "channel capacity" of the human operator on the total system performance is not known.

d. Systems with Random Parameters [23]. When a system has random parameters it is often difficult to describe the system response. The work of Foy [23] is an attempt to bridge this gap through the use of entropy. The class of systems investigated may be described by ordinary linear differential equations. The input forcing function and the coefficients of the differential equation have a known mathematical structure but contain certain parameters which have known probability distributions. The effect of the parameter variations on the output can then be measured through the use of the "instantaneous" output entropy (i.e., the entropy measure of uncertainty as to the value of the output that will be observed for a given value of the independent variable) as opposed to the use of the ensemble entropy (i.e., the entropy uncertainty as to which one of the class of output functions will be observed) for the same situation.

Unfortunately the general formula for the entropy of the output, that is in this dissertation, is mathematically intractable so

it is necessary for Foy to derive upper bounds, through the use of the properties of the Gaussian probability density function. The effectiveness of these bounds is supported by an experimental program on an analog computer. Examples, based on a useful and recognizable feedback loop system are presented but it is these examples that point out the fundamental weakness of the report, that weakness being that the evaluation of instantaneous output entropy has no useful interpretation, as yet. It does not appear to be a meaningful design criterion.

The reason for this may not be the inappropriateness of the entropy measure but the inconclusiveness of the research. This inconclusiveness may yet be resolved if Foy's work can be reevaluated in the context of an adaptive control system analysis. Certainly the usefulness of any adaptive controller must be measured and the results of current research clearly indicates the effectiveness of entropy for just this purpose. Foy's work may well prove to be the cornerstone of a totally new approach to adaptation.

Despite these examples described above, for the most part it has still been the motivation of the problems of communication, i.e., the problem of transmitting knowledge from point to point that has attracted the attention of the information theorists [24,25,26,27,28]. Reduced to its simplest abstraction, the communication problem concerns itself with; (1) a signal (or message) drawn in some fashion from a predefined signal vocabulary, (2) a channel for transmitting the message, and (3) a receiver for deciding the content of the received signal. By

assigning a measure to the information content of a message, information theory provides an invaluable tool for assessing the worth of the coding procedure, analyzing the performance of the channel, and describing the efficiency of the receiver. By tracing the message content from point to point in the communication system the concept of a system channel capacity follows directly and various design approaches evolve. In fact, the theory of channel capacity represents the major contribution of Shannon's pioneering effort in communication theory. With the fundamental channel capacity theorem given by

$$C = B \text{ LOG}_2 \left[1 + \frac{P}{N} \right] \text{ bits},$$

Shannon [1] showed conclusively how it was possible to exchange channel bandwidth (B) for signal power (P) in order to maintain constant channel capacity (C) for the same noise power (N).

The mathematical description of the channel behavior becomes more obscure when the signals involved are continuous. The process of passing from the discrete situation, with its secure foundation of intuitively acceptable results, to the continuous case is extremely tortuous, especially in view of the spectre of the ever-present paradox of a signal carrying infinite information in zero time [29,30]. This was a weak spot in the original Shannon work and to avoid his ambiguities and paradoxes it has been necessary for Gelfand [31], Pinsker [32], and Hyang [33] to each make new definitions. Hyang's work is probably the most up-to-date research on the entropy of time continuous processes and for background it contains a lengthy and important

discussion of the necessity for carefully defining a continuous-time-process information measure.

Hyang overcomes the most pressing difficulties in the following manner. Let

$$S_i = \int_0^T s(t)\phi_i(t)\,dt \qquad i = 1, 2, \ldots, K$$

$$Y_j = \int_0^T y(t)\,\mu_j(t)\,dt \qquad j = 1, \ldots, M$$

where

$$y(t) = s(t) + n(t).$$

$s(t)$ is the message and ϕ_i and μ_j are arbitrary functions. The average information in $y(t)$ about $s(t)$ is defined as

$$I(y(t);s(t)) = \sup I(Y_1, Y_2, \ldots, Y_M; S_1, \ldots, S_K)$$

where the supremum is taken over all ϕ_i, μ_j, K and M.

The powerfulness of this definition can be seen from the fact that there are no restrictions on

1. The process spectrum
2. The process stationarity
3. The observation interval.

It is therefore not surprising to learn that no general results have been obtained using this definition, so that Hyang then finds it convenient to consider only the class of Gaussian random processes. For this class of problems the covariance function describes completely the process probability density function, so that after some manipu-

lation it can be shown that the necessary basis functions for the information quantity are the eigenfunctions found from

$$\lambda_K \int_0^T R_N(t,u)\Phi_K(u)\, du = \sigma \int_0^T R_S(t,u)\Phi_K(u)\, du.$$

This integral equation is a fundamental equation in mean square error analysis and in that context it has led to the formulation of the Wiener-Hopf equation [34] for linear estimation and the formulation of polynomial estimators for nonlinear estimation [35]. Of course these classical theories now become even more interesting because they have been derived from a purely information theoretic point of view.

Hyang's results are important for two reasons; first they lead to valid results in the almost totally uninvestigated field of continuous process entropy and second, they indicate to even the casual observer that stronger (stronger than Gaussian) results for the solution of estimation problems may be obtained through the use of entropy analysis.

The very extensive and quite practically oriented examples chosen for demonstration of the application of the time continuous theory is only further proof that <u>information theory</u> has far exceeded <u>information practice</u>, because at no time is Hyang able to present a clear motivation for wanting to calculate continuous time entropy in the first place. In general, it is not the theoretical limitations of time continuous entropy that limits the usefulness of entropy techniques for describing and analyzing objects. Rather, it is the

assumed requirement for an optimal code which achieves arbitrarily small error for the cost of faithful reproduction of a message must, in general, be paid for with infinite time delays.

Such considerations and restrictions are unheard of in the design of feedback control systems. Conventional feedback control systems are usually continuous, long time delays are intolerable, and very rarely are there any coding considerations*. However, these differances between communication systems and feedback systems do not automatically preclude the application of information theory to control system analysis. Actually, it is in the areas of degradation of performance due to noise and channel capacity that communication and feedback systems show the most resemblances. The purpose of this dissertation is to show how the basic theory of information may be extended to take advantage of these similarities and accommodate the large class of feedback control problems.

If new results are to be obtained, it appears that it will be necessary in some way to rise above the restrictions implied by the coding approach to information theory and to reinterpret the concept of entropy in terms generic to the operation of a feedback control system as opposed to the operation of a communication system. For example, by abandoning coding concepts, Weiner [36] is able to use mutual information to derive filtering equations that are usually

*When a coding problem does arise, such as in a system using analog and digital components in a hybrid configuration, the coding is not germane to the feedback problem.

arrived at by the conventional methods of mean square error analysis. In the field of statistical analysis, Gardner and McGill [37] are able to demonstrate an entropy approach to the partitioning of variability, while Chaitanya Swarup [38] develops an informational description of the properties of estimators and hypothesis tests. Foy's work, already cited [23], is still another example of the non-coding aspects of entropy being applied for analysis purposes.

Krasovskiy [35,40], who considers a system that may be described by a set of nonlinear differential equations with time varying non-random coefficients, takes a slightly different point of view than that taken by Foy for investigating the entropy of the output of a dynamical system. When the unforced dynamical system is linear and can be described as

$$\frac{dx_i}{dt} + \sum_{K=1}^{N} a_{iK}(t)\, x_K = 0 \qquad (i=1, 2, \ldots, N).$$

Then he shows that the instantaneous entropy of the state vector,

$$\underline{X} = (x_1, x_2, \ldots, x_n)$$

may be found from

$$\frac{dH}{dt} = - \sum_{i=1}^{N} a_{ii}(t).$$

A derivation is also presented by Krasovskiy which allows for a random forcing function. Unfortunately this formulation is too unwieldy to be of use for anything but the Gaussian situation. In addition to this fault the paper suffers from two other important defects:

1. The equations of motions must be written for the system as a whole; subsystems can not be examined separately and then connected output-to-input. If the result of the entropy analysis is to be used to design a system component, then the form of the differential equation description of that component must be known down to the last parameter.

2. Krasovskiy presents no motivation for calculating the output entropy. Examples are presented and this quantity is determined but there is no justification for doing so.

Because the techniques of Foy and Krasovskiy begin with the system equations of motion, both are general enough to include feedback systems. But once the dynamical equations are written, it is no longer apparent whether the original system did or did not have a feedback path. Certainly it should be expected that any useful application of entropy as an analysis tool should maintain the distinction between feedback and non-feedback systems; and, in some cases, even provide greater insight into the basic differences between them.

The ability of the system to act upon a command message and to control, through the use of a feedback mechanism, an external object, most clearly distinguishes the feedback system from the communication system, and yet this has always been the most overlooked aspect of the problem. Even in Weiner's Cybernetics [36], which appears to be the earliest attempt to make use of entropy as a statistical design property, this fundamental distinction between the control problem and the filtering problem is ignored. Yet surprisingly, the reward

for insisting that the control system equations specifically emphasize the feedback form of the configuration is that when an entropy analysis is made the resulting parameters that describe the system performance can all be calculated from knowledge of the open loop behavior of the system.

The development and proof of this very general result represents the main contribution of this dissertation.

1.2 General Description of Problems Considered in this Dissertation

Two classes of problems are considered in this dissertation. They are:

1. The estimation problem shown in block diagram form in Figure 1.1, and
2. the feedback problem* shown in Figure 1.2.

Even though they represent fundamentally different applications, it is nevertheless true that both problems have a great deal of similarity. Both utilize noisy measurements in order to obtain a quantity which, when subtracted from the processed signal, produces an error which is minimum in the sense of a predefined criterion. In fact it is the demonstration of the similarity of the behavior of the noisy sensor in both types of systems that unifies these problems and which occupies the major effort of the analysis contained in this report.

*In the category of feedback control problems is included both noise rejecting and tracking systems. Where it is appropriate to do so, both types of systems will be referred to simply as feedback systems.

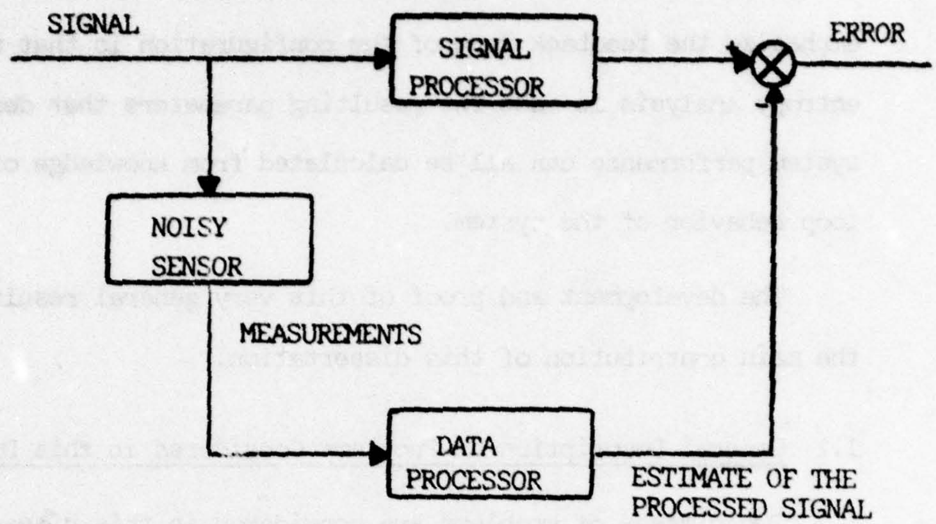


Figure 1.1. The Estimation Problem.

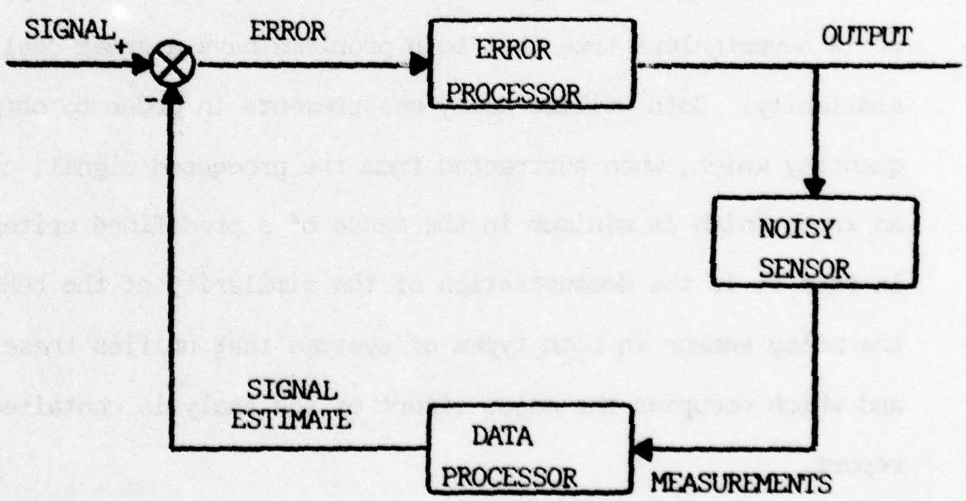


Figure 1.2. The Feedback Problem.

All the variables encountered in the problems in this dissertation are assumed to be sampled in time [not necessarily regularly], and to be continuous in an arbitrary but predefined probability space. The signal is taken to be statistically independent of the sensor errors.

Practical experience seems to indicate that the performance of these systems must be specifically limited by the behavior of the noisy sensor and yet conventional applications of presently accepted analysis techniques does not usually lead to this desired interpretation. In a very heuristic way a verbal description of the signal flow couched in the information theory language of uncertainty and channel capacity <u>seems</u> to provide the right approach toward evaluating the effect of the sensor on the performance attainable by such systems. However, even though the existance of a new theory of analysis based on sensor properties is suspected, a great deal of sophisticated manipulation involving entropy expressions is required in order to develop this theory, as will be seen later.

1.3 Brief Summary of Contents

Chapter Two develops the concept of entropy into a useful analysis tool. Beginning with a motivation for using entropy and related concepts to analyze control problems, it continues by outlining some of the more important formal properties of both entropy and mutual information and how these quantities change when acted upon by circuit elements. It is convenient to also introduce in this chapter, the concepts of entropy of markoff sources, entropy of sum

signals, and the transmittance and incremental transmittance of noisy sensors. Alone, these quantities have no intrinsic value, but their evaluation leads to proficiency with the entropy criterion function and paves the way for the useful interpretation of the important results relating to analysis of control systems which comes later. The channel transmittance of the sensor which is the information obtained about all of the signal samples from all of the measurements must, when the time comes to use the concept, be the limiting factor on the ability of a measurement path to improve system performance using either feed-forward or feedback signal processing. Markovian entropy on the other hand demonstrates that acquiring new measurements (or alternately saving old measurements) becomes less and less effective for reducing the signal uncertainty. This in turn leads to a proof of the existence of steady-state (non-perfect) performance solutions for the given systems.

Chapter Three initiates the effective use of entropy as a tool for the analysis of systems which estimate the values of random signals. The fundamental result is an equation relating the significant informational quantities encountered in this type of problem. This equation may then be rearranged to demonstrate:

1. Gaussian linear mean square error analysis is a special case of the entropy solution.

2. The error vector entropy for the system is bounded by a quantity which depends on the channel transmittance of the sensor.

3. The coordinate entropy of the error is bounded by a quantity which depends on the incremental channel transmittance of the sensor. These results bear two important resemblances to the pioneering work of Shannon. The first similarity, unfortunately, is a requirement that optimum system performance can only be obtained by delaying all data processing one entire message length. This is the "infinite delay" that is required by a Shannon code in the general case, in order to achieve error free transmission. For some types of estimation this is not a handicap, but real time estimation and feedback systems demand lagless data processing and can tolerate no delays. On the plus side, the second similarity is that, just as with the Shannon procedure, it is sufficient to prove the existence of a coding procedure [in the case of estimation to prove the existence of a filtering function] and evaluate the system performance as if the code were known. Never once, in the course of realizing the benefits of the work in this chapter is it actually necessary to know the optimizing filter.

These two observations of the theorem of Chapter Three hold out a promise of deriving a real time solution without sacrificing the decidedly important advantage of a "disappearing" filter function. Before this aspect of estimation is investigated, Chapters Four and Five are devoted to studying the feedback control problem and the tracking problem respectively. The results for both of these problems are very similar in form to the result derived for the estimation problem and are subject to the same sort of interpretations. Of course, as was suspected in the analysis of the conclusions to the theorem

of Chapter Three, the bounds derived for feedback systems are too loose and cannot be achieved through the use of physically realizable components. These results can not, therefore, be shown to be analogous to solutions derived by the application of Gaussian-linear mean square techniques. The important conclusions of these two chapters are:

1. Entropy analysis can lead to useful results.

2. Sensor channel transmittance is always a limiting factor to the system performance.

3. It is never necessary to know the optimum feedback filter or to evaluate the closed loop gain function, in order to determine the bounds on system entropy performance.

The real contribution of this dissertation is made in Chapter Six. Even though the channel approach to system theory is an original analysis technique, it is only when this theory is applied to real time data processing systems that significant advances in the understanding of feedback control systems are made. Chapter Six begins by deriving a real time property of sensors referred to as "The Sequential Channel Transmittance." This quantity relates the parameters of the sensor to its real time ability to provide new information about a signal when it is used in a suitable measuring path.

For both the estimation and feedback configurations it is proved that:

1. Gaussian linear mean square analysis of real time sampled data systems is a special case of the sequential entropy solution.

2. The entropy of the error is bounded by a quantity dependent on the sequential channel transmittance of the sensor.

3. All performance quantities can be determined without actually solving the optimum entropy problem for the solution filter.

Chapter Seven discusses the extension of the theories to continuous time systems. Based on an axiomatic description of uncertainty and mutual information, there is no reason to expect that all the results of the first six chapters do not carry over in total to the continuous time problem. Unfortunately, at the present time the theory of continuous entropy is not sufficiently developed to overcome the onus (or at least seemingly unexplainable fact) of infinite signal uncertainty. It is not known whether this limitation is due to the lack of experience on the part of researchers in understanding continuous entropy, or due to the need for an entirely new defining equations for entropy such as one based on integration in function space. In either case the example presented in this chapter does provide hope for the eventual development of satisfactory continuous time entropy analysis.

A very elementary approach to the problems of adaptive control is considered in Chapter Eight. Considering only one part of the adaptive process, the identification of unknown system parameters, an example and a theorem are presented which again demonstrate the importance of the sensor channel transmittance. While this is a new and

important result for the theory of adaptation it still remains to consider the entire adaptive process and to prove that the rate of improvement of the system performance is a function of the sensor channel properties.

Chapter Nine summarizes the principal results of the dissertation, discusses the limitations and disadvantages of the procedures developed and indicates the areas which require further investigation.

CHAPTER TWO
DEVELOPMENT OF ENTROPY AS AN ANALYSIS TOOL

2.1 Why Entropy?

Entropy (or uncertainty as it is often called) has found wide spread applications among communication engineers. When used to describe a process having a countable number of states, it has a very definite interpretation which leads from information content (bits) to coding length (bits/code symbol) to channel capacity (maximum bits/second) and then to practical system design. But its use is definitely limited when applied to continuous processes (even processes continuous in state space but discrete in time). Continuous signal entropy may be negative, mutual information can be infinite and even simple linear transformations are no longer entropy invariant. But these are not actually limitations on entropy as a system criterion function but more on entropy as an intuitively understandable concept. It is when entropy is treated as a cost function and not as measure of content of communication signals that it finds usage in the analysis of control systems.

The entropy H(x) of a scalar variable x, having a probability density function, $p_x(x)$ is defined as:

$$H(x) = \int_{-\infty}^{\infty} d\alpha \; p_x(\alpha) \log \frac{1}{p_x(\alpha)} \; .$$

In simple terms this function measures the "spread" of the random variable x. A variable whose pdf is more concentrated than another will have less entropy than the other variable. As a criterion,

entropy is analogous to the mean square error which measures the second order (the variance) spread of a variable. In fact, for Gaussian random variables there is a one-to-one relationship between variance and entropy so that when used as a criterion for system design minimum mean square error must always be equivalent to minimum entropy.

There are several other important probability density functions for which minimum entropy is equivalent to minimum mean square error. For example:

1. <u>The Rectangular Distribution</u>

$$p(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

$$E\{x\} = \frac{a+b}{2}$$

$$VAR\{x\} = \frac{(b-a)^2}{12} = \sigma^2$$

$$H(x) = LN\ (b-a) = LN\ \sigma + \frac{1}{2}\ LN\ 12$$

2. <u>The Exponential Distribution</u>

$$p(x) = \alpha e^{-\alpha x} \qquad x > 0$$

$$E\{x\} = \frac{1}{\alpha}$$

$$VAR\{x\} = \frac{1}{\alpha} = \sigma^2$$

$$H(x) = LN\ \frac{\varepsilon}{\alpha} = LN\ (\varepsilon\sigma^2)$$

This benign relationship between the variance and the entropy foretells great possibilities for the use of entropy as a tool for system performance analysis.

This relationship is, of course, not satisfied in general, so that minimum entropy designed systems are not always identical to

minimum mean square designed systems. However, it is not the use of entropy as a design tool that is considered in this report. Rather it is the use of entropy as a design criterion so as to set bounds on possible system performance and the inequalities relating entropy to variance that are considered. The ability to be able to write entropy expressions without first having to design the optimum system, distinguishes entropy analysis from conventional second order considerations.

The remainder of this chapter will be devoted to developing the basic properties of entropy and mutual information as they apply to systems analysis. Important theorems relating to:

1. The entropy of markoff sources,
2. the entropy of signals, conditioned on measurements,
3. the channel transmittance of a sensor, and,
4. the incremental channel transmittance of a sensor,

are presented in anticipation of a need for these results in the body of the dissertation. It is interesting to notice that the properties of entropy are in exact agreement with how intuition says an information measuring quantity should behave.

For the reader already familiar with the concepts of information theory, the results of the next two sections are summarized in Table I.

2.2 Properties of Entropy

The entropy of a K dimensional vector random variable having a probability density function that is continuous in all the components is defined by Shannon [1] on page 54 as:

TABLE I

Properties of Entropy and Mutual Information

1. $H(\underline{X}) \triangleq -\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \ldots \int_{-\infty}^{\infty} dx_K \; p_x(x_1,x_2,\ldots,x_K) \; \text{LOG}\; p_x(x_1,x_2,\ldots,x_K)$

   $= -\int_{-\infty}^{\infty} d\underline{X}\; p_x(\underline{X}) \;\text{LOG}\; p_x(\underline{X})$

2. $H(\underline{Y}) = H(\underline{X}) + \int_{-\infty}^{\infty} d\underline{X}\;\; p_x(\underline{X}) \;\text{LOG} \left| \text{DET} \left\{ \frac{\partial f_i(\underline{X})}{\partial x_j} \right\} \right|$

   where $\underline{Y} = F(\underline{X})$

3. $H(\underline{X}/\underline{Y}) \triangleq \int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{Y}\; p(\underline{X},\underline{Y}) \;\text{LOG}\; \frac{p_y(\underline{Y})}{p(\underline{X},\underline{Y})}$

4. $H(\underline{X},\underline{Y}) = \int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{Y}\; p(\underline{X},\underline{Y}) \;\text{LOG}\; \frac{1}{p(\underline{X},\underline{Y})}$

5. $H(\underline{Y},\underline{X}) = H(\underline{X},\underline{Y})$

6. $H(\underline{X},\underline{Y}) = H(\underline{X}) + H(\underline{Y}/\underline{X})$

7. $H(\underline{X},\underline{Y}) = H(\underline{Y}) + H(\underline{X}/\underline{Y})$

8. $H(Y) = H(y_1,y_2,\ldots,y_{K-1}) + H(y_K/y_{K-1},\ldots,y_2,y_1)$

9*. $H(\underline{X},\underline{Y}) \leq H(\underline{X}) + H(\underline{Y})$

10*. $H(\underline{X}/\underline{Y}) \leq H(\underline{X})$

11*. $H(\underline{Y}/\underline{X}) \leq H(\underline{Y})$

12. $H(\underline{Z}/\underline{Y},\underline{X}) \leq H(\underline{Z}/\underline{Y})$

13. $I(\underline{X};\underline{Y}) = \int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{Y}\; p(\underline{X},\underline{Y}) \;\text{LOG}\; \frac{p(\underline{X},\underline{Y})}{p_x(\underline{X})p_y(\underline{Y})}$

14*. $I(\underline{X};\underline{Y}) = I(\underline{Y},\underline{X}) \geq 0$

15. $I(\underline{X};\underline{Y}) = H(\underline{X}) + H(\underline{Y}) - H(\underline{X},\underline{Y})$

16. $I(\underline{X};\underline{Y}) = H(\underline{X}) - H(\underline{X}/\underline{Y}) = H(\underline{Y}) - H(\underline{Y}/\underline{X})$

17. $I(\underline{X};\underline{Y},\underline{Z}) \geq I(\underline{X},\underline{Y})$

***The equality holds if X and Y are independent random variables.**

$$H(\underline{X}) \triangleq - \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_K \; p_x(x_1,\ldots,x_K) \text{ LOG } p_x (x_1,\ldots,x_K)$$

$$= - \int_{-\infty}^{\infty} d\underline{X} \; p_x (\underline{X}) \text{ LOG } p_x(\underline{X}) \tag{2.2.1}$$

where the vector notation $\underline{X}$ = column $\{x_1,x_2,\ldots,x_K\}$ is used. In all cases in this dissertation capital letters are used to denote the vector and lower case letters the vector components.

From the communications point of view continuous entropy has three "disadvantages" when compared to the entropy of variables having discrete probability densities:

1. Continuous entropy is not always non-negative.
2. Continuous entropy is not always finite.
3. Continuous entropy is not always invariant under linear transformations.

These properties clearly handicap the interpretation of entropy as a measure of uncertainty. For control system applications, however, there is no obligation to view entropy as anything more than a suitable criterion function. As an example of the variability of entropy under coordinate transformations examine the vector random variables, $\underline{X}$, and $\underline{Y}$, where

$$\underline{Y} = F(\underline{X})$$

or

$$y_i = f_i(x_1,x_2,\ldots,x_K)$$

If $F(\underline{X})$ is continuous and one-to-one, the two probability density functions are related by (see Parzen [41], page 329-331):

$$p_x(\underline{X}) = p_y\ (F(\underline{X}))\ J$$

where J is the Jacobian of the transformation and it is defined as

$$J = \left| \mathrm{DET} \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_K} \\ \vdots & & \\ \dfrac{\partial f_K}{\partial x_1} & \cdots & \dfrac{\partial f_K}{\partial x_K} \end{bmatrix} \right| \triangleq \left| \mathrm{DET} \left\{ \frac{\partial f_i(\underline{X})}{\partial x_j} \right\} \right|$$

The entropy of $\underline{X}$ is then given by

$$H(\underline{X}) = -\int_{-\infty}^{\infty} d\underline{X}\ p_x(\underline{X})\ \mathrm{LOG}\ p_x\ (\underline{X})$$

$$= -\int_{-\infty}^{\infty} d\underline{X}\ p_y\ (F(\underline{X}))\ J\ [\mathrm{LOG}\ (p_y(F(\underline{X})) + \mathrm{LOG}\ (J)]$$

$$= -\int_{-\infty}^{\infty} dY\ p_y(\underline{Y})\ \mathrm{LOG}\ (p_y(\underline{Y})) - \int_{-\infty}^{\infty} d\underline{X}\ p_x(\underline{X})\ \mathrm{LOG}\ (J)$$

or finally

$$H(\underline{Y}) = H(\underline{X}) + \int_{-\infty}^{\infty} d\underline{X}\ p_x\ (\underline{X})\ \mathrm{LOG}\ (J) \qquad (2.2.2)$$

This result is due to Shannon [1], page 57, and is listed in Table I, as property 2. If $F(\underline{X})$ is the linear transformation

$$y = f(x) = ax + b, \qquad \text{with} \qquad J = a,$$

then the entropy of y is

$$H(y) = H(x) + \mathrm{LOG}\ |a|$$

Thus, depending on the value of "a," the entropy of y can either be greater or less than the entropy of x. H(y) = H(x) if and only if a = 1. The whole question of the entropy of transformed vectors is so important that it will be studied again in Section 2.4.

Associated with the pair of vectors $\underline{X}$ and $\underline{Y}$, which possess continuous marginal densities and a continuous joint probability density function, are the conditional entropies $H(\underline{X}/\underline{Y})$, $H(\underline{Y}/\underline{X})$ and the joint entropy $H(\underline{X},\underline{Y})$. These quantities are defined as:

$$H(\underline{X}/\underline{Y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\underline{X}\, d\underline{Y}\, p(\underline{X}, \underline{Y}) \text{ LOG } \frac{p_y(\underline{Y})}{p(\underline{X},\underline{Y})} \tag{2.2.3}$$

$$H(\underline{Y}/\underline{X}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\underline{X}\, d\underline{Y}\, p(\underline{X}, \underline{Y}) \text{ LOG } \frac{p_x(\underline{X})}{p(\underline{X},\underline{Y})} \tag{2.2.4}$$

$$H(\underline{X}, \underline{Y}) = \int_{-\infty}^{\infty} d\underline{X}\, d\underline{Y}\, p(\underline{X}, \underline{Y}) \text{ LOG } \frac{1}{p(\underline{X},\underline{Y})} \tag{2.2.5}$$

$$H(\underline{Y}, \underline{X}) = H(\underline{X}, \underline{Y}) \tag{2.2.6}$$

The following equations relate the entropy expressions:

$$H(\underline{X}, \underline{Y}) = H(\underline{X}) + H(\underline{Y}/\underline{X}) \tag{2.2.7}$$

$$= H(\underline{Y}) + H(\underline{X}/\underline{Y}) \;. \tag{2.2.8}$$

When $\underline{X}$ and $\underline{Y}$ are independent random variables the joint probability density function is simply

$$p(\underline{X}, \underline{Y}) = p_x(\underline{X})\, p_y(\underline{Y}),$$

and then

$$H(\underline{Y}/\underline{X}) = H(\underline{Y}) \tag{2.2.9}$$

$$H(\underline{X}/\underline{Y}) = H(\underline{X}) \tag{2.2.10}$$

$$H(\underline{X}, \underline{Y}) = H(\underline{Y}) + H(\underline{X}) \;. \tag{2.2.11}$$

The conditional entropy provides a simple way for expressing the entropy of a K dimensional vector as the sum of entropies of one dimensional entropies. First note that

$$p(x_1,x_2,\ldots,x_K)=p(x_K/x_{K-1},\ldots,x_1)p(x_{K-1}/x_{K-2},\ldots,x_1)\ldots p(x_2/x_1)p(x_1).$$

Then

$$H(\underline{X}) = H(x_K/x_{K-1},\ldots,x_2,x_1) + H(x_{K-1}/x_{K-2},\ldots,x_2,x_1)$$
$$+ \ldots + H(x_2/x_1) + H(x_1) . \qquad (2.2.12)$$

When the $\{x_K\}$ form a M^{th} order stationary markoff chain the conditional probability densities can be simplified using:

$$p(x_K/x_{K-1},x_{K-2},\ldots,x_2,x_1) = p(x_K/x_{K-1},\ldots,x_{K-M-1}) \qquad K \geq M .$$

For such a situation, $H(x_K/x_{K-1},\ldots,x_1) = H(x_K/x_{K-1},\ldots,x_{K-M-1})$ and

$$H(\underline{X}) = H(x_K/x_{K-1},\ldots,x_{K-1-M}) + H(x_{K-1}/x_{K-2},\ldots,x_{K-2-M})$$
$$+ \ldots + H(x_1,x_2,\ldots,x_M)$$
$$= (K-M)\, H(x_{M+1}/x_M,\ldots,x_1) + H(x_1,x_2,\ldots,x_M) , \qquad (2.2.13)$$

where the stationary property of the random sequence, x_1, x_2, x_3, $x_4,\ldots,$ x_K, is used to write this last equation.

For first order markoff components the total entropy of the vector $\underline{X}$ is

$$H(\underline{X}) = (K-1)\, H(x_2/x_1) + H(x_1) . \qquad (2.2.14)$$

Another useful property of the entropy of first order markoff process is that

$$H(\phi(x_k)/x_{k-1}, x_{k-2}) = H(\phi(x_k)/x_{k-1}) \qquad (2.2.15)$$

where

$$y_K = \phi(x_K) \quad k = 1, 2, \ldots$$

is an arbitrary single valued function of x, having a continuous (and non-zero) derivative.

If $p_1(x_K, x_{K-1}, x_{K-2})$ is the joint probability density function of x_K, x_{K-1}, and x_{K-2}, and if $p_2(y_K, x_{K-1}, x_{K-2})$ is the joint probability density function for y_K, x_{K-1}, and x_{K-2}, it is true that

$$p_2(y_K, x_{K-1}, x_{K-2}) = p_2(\phi(x_K), x_{K-1}, x_{K-2})$$
$$= p_1(x_K, x_{K-1}, x_{K-2}) \left| \frac{d\phi(x_K)}{dx_K} \right|^{-1} .$$

Then

$$p_2(y_K/x_{K-1}, x_{K-2}) = p_1(x_K/x_{K-1}) \left| \frac{d\phi(x_K)}{dx} \right|^{-1}$$

$$= \frac{p_1(x_K, x_{K-1})}{p(x_{K-1})} \left| \frac{d\phi(x_K)}{dx} \right|^{-1} = p_2(y_K/x_{K-1}) = p_2(\phi(x_K)/x_{K-1})$$

and equation (2.2.15) follows directly.

When the components of a vector are Gaussian random variables the probability distribution of the vector can be written as:

$$p(\underline{X}) = \frac{1}{(2\pi)^{K/2}\sqrt{\text{DET } R_x}} \; e^{-\frac{1}{2}\underline{X}^T R_x^{-1} \underline{X}} \tag{2.2.16}$$

where R_x is the covariance matrix with entries, $R_{ij} = E\{x_i, x_j\}$. The total entropy of the vector $\underline{X}$ is then

$$H(\underline{X}) = \int_{-\infty}^{\infty} d\underline{X} p(\underline{X}) \left[\frac{1}{2} \text{LOG } (2\pi)^K \text{ DET } R_x + \frac{1}{2} \underline{X}^T R_x^{-1} \underline{X} \right]$$

$$= \frac{1}{2} \text{LOG } (2\pi)^K e + \frac{1}{2} \text{LOG } [\text{DET } R_x] . \tag{2.2.17}$$

For two one-dimensional random variables, Shannon [1], has shown that for the same variance, the random variable with a Gaussian distribution always has a greater entropy than the random variable with any other distribution. This conclusion leads directly to the important inequality:

$$\mathrm{VAR}\,\{x\} \geq \frac{1}{2\pi e}\,\mathrm{LOG}\;2H(x) \tag{2.2.18}$$

where x has an arbitrary probability density function. This inequality provides the bridge between conventional second order analysis and entropy analysis.

A property of entropy which is extremely valuable is that as more data becomes available about the primary random variable the entropy of that random variable decreases. The proof of this statement concerning the conditional entropy of $\underline{Z}$ is embodied in the following theorem.

Theorem 2.2.1:

The entropy of $\underline{Z}$ decreases as it is conditioned on more data, i.e.,

$$H(\underline{Z}/\underline{Y}) \geq H(\underline{Z}/\underline{Y},\underline{X}) \;. \tag{2.2.19}$$

Proof:

$$H(\underline{Z}/\underline{Y}) - H(\underline{Z}/\underline{Y},\underline{X}) = \int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{Z}\; p(\underline{X},\underline{Y},\underline{Z})\;\mathrm{LOG}\, \frac{\dfrac{p(\underline{X},\underline{Y},\underline{Z})}{p(\underline{Y},\underline{X})}}{\dfrac{p(\underline{Y},\underline{Z})}{p(\underline{Y})}}$$

$$= \int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{Z}\; p(\underline{X},\underline{Y},\underline{Z})\;\mathrm{LOG}\, \frac{p(\underline{X},\underline{Y},\underline{Z})}{p(\underline{Y},\underline{Z})\,\dfrac{p(\underline{Y},\underline{X})}{p(\underline{Y})}}$$

Now use the inequality

$$\text{LN}\,\frac{1}{\alpha} \geq 1-\alpha$$

to get

$$H(\underline{Z}/\underline{Y}) - H(\underline{Z}/\underline{Y},\underline{X}) \geq \int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{Z}\; p(\underline{X},\underline{Y},\underline{Z}) \left[1 - \frac{p(\underline{Y},\underline{Z})\; p(\underline{Y},\underline{X})}{p(\underline{X},\underline{Y},\underline{Z})\; p(\underline{Y})} \right]$$

$$\geq 1 - \int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{Z}\; \frac{p(\underline{Y},\underline{Z})\; p(\underline{Y},\underline{X})}{p(\underline{Y})}$$

$$= 1 - \int_{-\infty}^{\infty} d\underline{Y}\; \frac{p(\underline{Y})\; p(\underline{Y})}{p(\underline{Y})} = 1-1 = 0$$

$$H(\underline{Z}/\underline{Y}) - H(\underline{Z}/\underline{Y},\underline{X}) \geq 0 \qquad \text{Q.E.D.}$$

This theorem is an obvious extension of a similar theorem given by Shannon [1] for the scalar case, i.e., by Shannon:

$$H(z) \geq H(z/y)$$

The technique for the proof is an accepted procedure for proving entropy inequalities and has been most effectively used by Feinstein [2] and Abramson [7].

Since much of the solution to the estimation and feedback control problems involves calculating the uncertainty (e.g., entropy) of the signal sample given the measurements made of the signal vector, it is important to consider further the properties of the entropy of conditioned signals. However, even before the actual need for this function arises (it first appears in 3.7) it is obvious that it must be part of any effort to determine the effectiveness of a

procedure to predict a signal based on noisy measurements of that signal.

It is expected that the uncertainty of the signal will decrease as more measurements are made. However, it is not expected that uncertainty is likely to approach certainty as the number of measurements approaches infinity. While this is a minor point, it spells the difference between understanding the steady-state behavior of a system or being completely ignorant of any limiting factors to its performance. The important attributes of the conditional entropy function, are proven in the following theorem. The inspiration for this theorem and the procedure are due to Birch [15] but the statement, proof and interpretation are original.

Theorem 2.2.2:

For stationary processes the conditional entropy of the scalar random variable y_K, given the last K values of the noise corrupted signal, is a monotonically decreasing function of K and has a finite limit, i.e.,

$$h_K \stackrel{\Delta}{=} H(y_K/\underline{Z}_K) \geq H(y_{K+1}/\underline{Z}_{K+1}) = h_{K+1} \qquad (2.2.20)$$

and

$$\lim_{K\to\infty} H(y_K/\underline{Z}_K) = H(y_K/\underline{Z}_\infty) > -\infty \qquad (2.2.21)$$

Proof:

The monotonic behavior of the conditional entropy follows directly from the stationarity of the processes and the previous theorem (2.2.1).

$$H(y_K/\underline{Z}_K) = H(y_{K+1}/z_{K+1}, z_K, \ldots, z_2)$$

$$\geq H(y_{K+1}/z_{K+1}, \ldots, z_2, z_1) = H(y_{K+1}/\underline{Z}_{K+1})$$

Thus, the entropy of y_K, conditioned on all the past measurements is a decreasing function of K, where $\underline{Z}_K$ is the vector whose components, $z_k = y_k + n_k$, are the K noise measurements made of y_k, $k = 1, 2, \ldots, K$.

It is inconceivable that in ordinary physical situations $H(y_K/\underline{Z}_K) = -\infty$, for any finite K*. Singular probability distributions can be imagined where a finite number of measurements can be used to predict the signal with probability one [entropy of $-\infty$], but such situations are of no real concern. The real concern is to be able to bound $H(y_K/\underline{Z}_K)$ away from $-\infty$. Obviously additional conditions must be imposed to insure that with even an infinite number of measurements the conditional uncertainty of the signal is not $-\infty$. Consider first the situation when y is first order markoff, the extension to M^{th} order processes will be obvious.

It is convenient at this time to introduce another sequence $\{g_K\}$ defined by:

$$g_{K+1} \stackrel{\Delta}{=} H(y_{K+1}/\underline{Z}_{K+1}, y_1) .$$

Because of the process stationarity it is also true that

$$g_{K+1} = H(y_K/z_K, z_{K-1}, \ldots, z_o, y_o).$$

---

*An entropy of $-\infty$ implies a completely certain continuous random variable.

The sequence $\{g_K\}$ is now shown to be a monotonically increasing function of K

$$g_{K+1} = H(y_{K+1}/z_1,z_2,\ldots,z_{K+1},y_1)$$
$$= H(y_{K+1}/z_1,z_2,\ldots,z_{K+1},y_1,n_1),$$

where the rearrangement of the pair (z_1,y_1) into (z_1,y_1,n_1) through the use of $z_1 = y_1 + n_1$, does not change the conditional entropy of y_{K+1}. Since the $\{y_K\}$ and $\{n_K\}$ are taken as first order markoff sequences, the result of equation (2.2.15) implies that the introduction of y_o and n_o into the expression for g_{K+1} will not change the forward conditional probabilities, i.e.,

$$g_{K+1} = H(y_{K+1}/z_1,z_2,\ldots,z_{K+1},y_1,n_1,y_o,n_o)$$
$$= H(y_{K+1}/z_o,z_1,\ldots,z_{K+1},y_1,y_o)$$
$$\leq H(y_{K+1}/z_o,\ldots,z_{K+1},y_o) \triangleq g_{K+2},$$

thus proving that g_K is a monotonically increasing function of K. Now, the following inequality describes the relationship between the sequences $\{h_K\}$ and $\{g_K\}$. For all $k \geq \kappa$

$$g_\kappa \leq g_K \leq h_K \leq h_\kappa .$$

Obviously both sequences are bounded and monotonic so that the following limits exist:

$$\lim_{K\to\infty} h_K = \overline{h}$$

$$\lim_{K\to\infty} g_K = \overline{g}$$

and

$$g_K \leq \overline{g} \leq \overline{h} \leq h_K \qquad \forall K .$$

It follows that $\lim H(y_K/\underline{Z}_K)$ is bounded away from $-\infty$, if g_K is not $-\infty$ for some finite K. Examine g_2.

$$g_2 = H(y_2/z_1,z_2,y_1),$$

when the $\underline{Z}$ process is formed by adding independent random variables, i.e.,

$$z_k = y_k + n_k \qquad k = 0, 1, \ldots, K$$

it is not possible that

$$g_2 = -\infty$$

and the theorem is proven.

2.3 Mutual Information

Communication theory has given rise to a quantity, which because of its properties, is even more valuable than entropy. This quantity is "mutual information." Technically, the mutual information, $I(\underline{X};\underline{Y})$, between the two random vectors $\underline{X}$ and $\underline{Y}$ is

$$I(\underline{X};\underline{Y}) = I(\underline{Y};\underline{X}) = H(\underline{X}) - H(\underline{X}/\underline{Y})$$

$$\triangleq -\iint d\underline{X}\, d\underline{Y}\, p(\underline{X},\underline{Y}) \operatorname{LOG} \frac{p(\underline{X})p(\underline{Y})}{p(\underline{X},\underline{Y})} . \qquad (2.3.1)$$

Unless otherwise noted, both $\underline{X}$ and $\underline{Y}$ are usually taken as K dimensional vectors.

If $H(\underline{X})$ is the a priori entropy of $\underline{X}$ and $H(\underline{X}/\underline{Y})$ is the entropy of $\underline{X}$ after observing $\underline{Y}$, then $I(\underline{X};\underline{Y})$ is the average amount of entropy supplied by $\underline{Y}$. However, it is often convenient to use the intuitive notion that $I(\underline{X},\underline{Y})$ is the average amount of "information" obtained about $\underline{X}$ by being given the value of $\underline{Y}$.

Mutual information (or transinformation) is

1. symmetrical in $\underline{X}$ and $\underline{Y}$,
2. non-negative ($I(\underline{X};\underline{Y}) \geq 0$),
3. generally finite,
4. invariant under linear transformations.

In addition, the following relationships are satisfied:

1. $I(\underline{X};\underline{Y}) = 0$ (if $\underline{X}$ and $\underline{Y}$ are independent)
2. $I(\underline{X};\underline{Y}) = H(\underline{X}) + H(\underline{Y}) - H(\underline{X},\underline{Y})$ (2.3.2)
3. $I(\underline{X};\underline{Y}) = H(\underline{X}) - H(\underline{X},\underline{Y})$ (2.3.3)
4. $I(\underline{X};\underline{Y}) = H(\underline{Y}) - H(\underline{Y}/\underline{X})$ (2.3.4)

Using $I(\underline{X};\underline{Y}) \geq 0$ in equations (2.3.2), (2.3.3) and (2.3.4), yields

$$H(\underline{X}) + H(\underline{Y}) \geq H(\underline{X},\underline{Y}) \tag{2.3.2a}$$

$$H(\underline{X}) \geq H(\underline{X}/\underline{Y}) \tag{2.3.3a}$$

$$H(\underline{Y}) \geq H(\underline{Y}/\underline{X}) \tag{2.3.4a}$$

With mutual information just inversely as with entropy, acquiring more measurements increases mutual information montonically. This is stated as:

Theorem 2.3.1:

If the information about $\underline{X}$ given $\underline{Y}$ is $I(\underline{X};\underline{Y})$, then the information about $\underline{X}$ given $\underline{Y}$ and $\underline{Z}$ must be greater, i.e.,

$$I(\underline{X};\underline{Y}) \leq I(\underline{X};\underline{Y},\underline{Z}) \tag{2.3.5}$$

where the coordinates of $\underline{Z}$ are additional coordinates of the $\underline{Y}$ vector (or are other observations).

Proof:

The statement for this theorem is given by Gel'Fand [31], but his proof is much too sophisticated for the analysis of sample data

signals so a simpler one is supplied here. Using the definition of mutual information, $I(\underline{X};\underline{Y},\underline{Z})$ is

$$I(\underline{X};\underline{Y},\underline{Z}) = -\int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{Z}\; p(\underline{X},\underline{Y},\underline{Z}) \text{ LOG } \frac{p(\underline{X})p(\underline{Y},\underline{Z})}{p(\underline{X},\underline{Y},\underline{Z})}$$

$$= -\int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{Z}\; p(\underline{X},\underline{Y},\underline{Z}) \text{ LOG } \frac{p(\underline{X})p(\underline{Y})p(\underline{Z}/\underline{Y})}{p(\underline{X},\underline{Y})p(\underline{Z}/\underline{X},\underline{Y})}$$

$$I(\underline{X};\underline{Y},\underline{Z}) = I(\underline{X};\underline{Y}) + H(\underline{Z}/\underline{Y}) - H(\underline{Z}/\underline{X},\underline{Y})$$

and finally, using theorem 2.2.1, this equation becomes

$I(\underline{Z};\underline{Y},\underline{Z}) \geq I(\underline{X};\underline{Y})$. Q.E.D.

2.4 Coordinate Transformations

Much of the work of this dissertation depends on considering the entropy (or mutual information) of transformed variables. It is therefore convenient for the initial developments to restrict the derivations to the class of information preserving transformations.

If the mutual information between a pair of random variables is invariant under a transformation of those variables it is not intuitively obvious that the given transformation must be one-to-one, or that the inverse transformation exists so that the input vector can always be "recovered" if the output vector is known. When studying random variables with continuous (or at least piecewise continuous) probability density functions the constraint on the transformation is made slightly more severe than merely being

information preserving, so as to include the "recovery" property. The following definition will accomplish this.

<u>Definition:</u>

An information preserving transformation from $x \xrightarrow{f(x)} y$ is one for which the equation defining the transformation of the probability density functions [41],

$$p_x(x) = p_y(f(x)) \left| \frac{df}{dx}(x) \right| , \tag{2.4.1}$$

is defined and is valid for all values of x.

It is quite easy to show that when f(x) satisfies the conditions of equation (2.4.1) it is a regular transformation, i.e., both f(x) and $f^{-1}(y)$ are of class $C^{(1)}$ and $I(x;z) = I(f(x);z) = I(y;z)$.

Alternately an equivalent constraint on f(x) is that the Jacobian of the transformation is continuous and non-zero. In any case, the important fact is that when the conditions on f(x) are defined, in either manner, f(x) has an inverse and no information about the input is lost because of the transformation.

Actually, the true class of information invariant transformations is much more inclusive than the class of functions defined by the use of equation (2.4.1). This is because the broader class of transformations also includes functions, $F(\cdot)$, that while not having a unique inverse, do possess the property that for all sets A and B, where $B = F^{-1}(A)$ it is true that

$$p(z/y\varepsilon A) = p(z/x_1\varepsilon B) = p(z/x_2\varepsilon A) \qquad \forall x_1,x_2\varepsilon B \quad . \tag{2.4.2}$$

Since $I(z;y) = H(z) - H(z/y)$, it must follow that for any transformation

$$y = f(x)$$

satisfying equation (2.4.2)

$$I(z;x) = I(z;f(x)) = I(z;y).$$

As an example of this type of function, consider

$$y = x^2$$

together with

$$p(z/x) = 0 \qquad x \leq 0.$$

Now let R be the continuous real line and A and B be the positive half line (including zero). Then obviously

$$p(z/y \epsilon A) = p(z/x \epsilon B)$$

and therefore, even though $f^{-1}(y)$ is not unique,

$$I(z;x) = I(z;y).$$

This is a "contrived" example and it is not likely that transformations matching themselves so closely to the conditional density $p(z/x)$ so as to satisfy equation (2.4.2) will ever occur in practical situations.

2.5 Entropy of Markoff Sources

In general, most of the random processes encountered in control systems analysis are markoff sources which may be considered as being generated by the simple M^{th} order differencing process given by:

$$y_1 = \xi_1$$

$$y_2 = a_{21}y_1 + b_{21}\xi_1 + \xi_2$$

$$\vdots$$

$$y_k = \sum_{i=k-M}^{M} a_{ki}y_i + \sum_{i=k-M}^{M} b_{ki}\xi_i \qquad k>m$$

$$\vdots$$

$$y_K = \sum_{i=K-M}^{M} a_{Ki}y_i + \sum_{i=K-M}^{M} b_{Ki}\xi_i \tag{2.5.1}$$

where the driving function $\{\xi_k\}$ is a sequence of independent random variables chosen with the necessary probability density functions to achieve the desired density for the y sequence. It is not necessary that the ξ process be stationary or that the differencing be time invariant. However, later on such simplifications will be introduced in order to study the much more common stationary markoff process.

For convenience of notation it is desirable to make the following definitions:

$$\underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \\ \vdots \\ y_K \end{bmatrix}, \qquad \underline{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_k \\ \vdots \\ \xi_K \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \; . \; . \; . & 0 \; . \; . \; . \\ (a_{21}+b_{21}) & 1 & 0 \; . \; . \; . \\ \begin{bmatrix} a_{32}a_{21}+a_{32}b_{21} \\ +a_{31}+b_{31} \end{bmatrix} & a_{32}+b_{32} & 1 \; . \; . \; . \\ \vdots & \vdots & \vdots \end{bmatrix}$$

A is a lower triangular matrix, i.e., it is the matrix of a physically realizable process, and its determinant is 1. Using a vector algebra the generation of $\{y_k\}$ can be written simple as:

$$\underline{Y} = A\underline{\xi} \qquad (2.5.2)$$

The entropy of the $\underline{Y}$ vector can now be determined directly from the entropy of the generating sequence $\underline{\xi}$ using the relationship

$$H_K(\underline{Y}) = H_K(\underline{\xi}) + \text{LOG}\ [\text{DET}\ (A)] = H_K\ (\underline{\xi})$$

where the subscript K is used to denote that the dimension of the vector whose entropy is being described is K.

Since $\underline{\xi}$ is composed of independent random variables it follows directly that

$$H_K\ (\underline{Y}) = \sum_{k=1}^{K} H_1(\xi_k)\ . \qquad (2.5.3)$$

Now if an additional value of y, namely y_{K+1}, is generated then

$$H_{K+1}(\underline{Y}) = \sum_{k=1}^{K+1} H_1(\xi_k) = \sum_{k=1}^{K} H_1(\xi_k) + H_1(\xi_{K+1})$$

$$= H_K(\underline{Y}) + H_1(\xi_{K+1}) \tag{2.5.4}$$

but

$$H_{K+1}(\underline{Y}) = H_K(\underline{Y}) + H_1(y_{K+1}/y_K, y_{K-1}, \ldots y_1). \tag{2.5.5}$$

Therefore, these last two expressions together with the markoff property of the y process yields

$$H_1(y_{K+1}/y_K, \ldots, y_{K-M}) = H_1(\xi_{K+1}) \qquad K>M. \tag{2.5.6}$$

When written in terms of the probability density functions this equation is:

$$\int_{-\infty}^{\infty} dy_{K+1} p(y_{K+1}, y_K, y_{K-1}, \ldots, y_{K-M}) \text{ LOG } \frac{p(y_K, \ldots, y_{K-M})}{p(y_{K+1}, y_K, \ldots, y_{K-M})}$$

$$= \int_{-\infty}^{\infty} d\xi p_\xi(\xi_{K+1}) \text{ LOG } \frac{1}{p_\xi(\xi_{K+1})} .$$

In terms of the initial entropy of the y sequence, and the conditional increase in entropy for each additional y generated, the total entropy of the $\underline{Y}$ vector is

$$H_K(\underline{Y}) = \sum_{k=M+1}^{K=1} H_1(y_{k+1}/y_k, \ldots, y_{k-M}) + H_M(y_1, \ldots, y_M) . \tag{2.5.7}$$

Define the entropy of an M^{th} order markoff source as

$$H^*(y_{k+1}/\underline{Y}) \triangleq H_1(y_{k+1}/y_k, \ldots, y_{k-M}) . \tag{2.5.8}$$

Then for the case when the probability distribution of ξ_k is stationary:

$$H_K(\underline{Y}) = \sum_{k=M+1}^{K-1} H^*(y_{K+1}/\underline{Y}) + H_M(y_1,\ldots,y_M) = KH(\xi) . \qquad (2.5.9)$$

If the differencing process is time invariant, i.e., if $b_{ij}=b_{ik}$, and $a_{ij}=a_{ik}$ for all i, j, k>M, then

$$H_K(\underline{Y}) = (K-M)\, H^*(y_{K-1}/\underline{Y}) + H_M(y_1,\ldots,y_M) . \qquad (2.5.10)$$

When the driving function $\underline{\xi}$ is Gaussian the $\underline{Y}$ vector is also Gaussian, so that the total entropy of the y process is

$$H_K(\underline{Y}) = \frac{1}{2} \text{LOG } (2\pi)^K e\Delta_Y \qquad (2.5.11)$$

where

$$\Delta_Y = \text{DET } [E\{\underline{YY}^T\}] = \text{DET } [R_{yy}]$$

$$R_{yy} = E\,\{\underline{YY}^T\} = E\,\{A\underline{\xi\xi}^T A^T\} = AR_{\xi\xi}\, A^T$$

$$R_{\xi\xi} = \begin{bmatrix} \sigma_{\xi_1}^2 & & & & & \\ & \sigma_{\xi_2}^2 & & & & \\ & & \sigma_{\xi_3}^2 & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \cdot \end{bmatrix}$$

so finally

$$\Delta_Y = \text{DET } [AR_{\xi\xi}A^T] = \text{DET } [A]\ \text{DET } [R_{\xi\xi}]\ \text{DET } [A^T] \qquad (2.5.12)$$

$$= \prod_{k=1}^{K} \sigma_{\xi_k}^2 .$$

Then

$$H_K(\underline{Y}) = \sum_{k=1}^{K} \frac{1}{2} \text{LOG } 2\pi e\ \sigma^2_{\xi_k} = \sum_{k=1}^{K} H_1(\xi_k) \tag{2.5.13}$$

which is, of course, identical to equation (2.5.3).

As before, for the first order markoff source:

$$H_1(y_K/y_{K-1}) = H_1(\xi_K)\ . \tag{2.5.14}$$

When the ξ_k distribution is stationary and the differencing process is time invariant, the conditional entropy for very large K is found from

$$E\ \{y_K y_{K-1}\} = \sigma_Y^2\ a \tag{2.5.15}$$

and

$$\lim_{K\to\infty} E\{y_K y_{K-1}\} = \sigma_\xi^2 \frac{a}{1-a^2}$$

so that

$$H_1(y_K/y_{K-1}) = \frac{1}{2}\ \text{LOG } 2\pi e\ \sigma_\xi^2 = \frac{1}{2}\ \text{LOG } 2\pi e\ \sigma_y^2\ (1-a^2) \tag{2.5.16}$$

2.6 The Channel Capacity of a Sensor

All real world systems must make use of one or more measuring devices in order to be able to have an absolute (or relative) assessment of performance so that the tasks which are performed may be judged. It seems intuitively true that in any real time control problem, feed-forward or feedback, it must be the measuring device which ultimately limits the performance of the control process. If data could be taken quickly, efficiently and accurately, then 100 percent precise commands could be given and zero error performance

would be achieved. Of course no sensor can provide data that good; they all provide data that is distorted, noisy, late and sometimes missing. Until this present report, feedback control system theory has lacked techniques for assessing the worth of sensors. There has been no good criterion by which sensors could be judged and no scale on which alternate sensors could be compared.

The significant new results in this dissertation were obtained by no longer using the sensor as a data gathering device having certain variance properties, but rather as an information transmitting device having channel properties. Such a view is not unnatural. It is a highly respectable assumption when made by coding theorists analyzing communication systems and moreover much can be said for the analogy between control and communication.

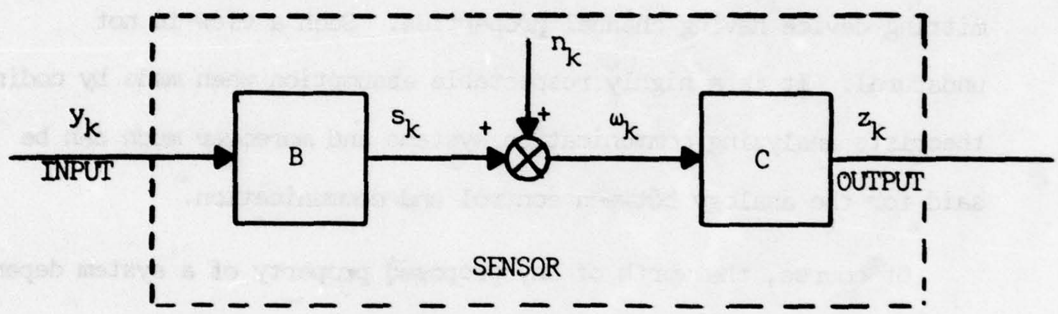
Of course, the worth of any proposed property of a system depends on the knowledge it conveys to the user, and given that a sensor has a channel property it is not surprising to learn (as will be proven in the later chapters) that this property can be used to develop a measure of the system performance. The channel capacity of the sensor becomes the factor which limits the flow of information around a system and therefore the factor which can measure the effectiveness of the given sensor path in accomplishing some end.

The channel property studied in this dissertation will be that of the mutual information between sensor input and output and will be referred to as the "Sensor Channel Transmittance." This quantity closely resembles Shannon's channel capacity, but, for the

situations examined here, the luxury of a controllable signal probability density function is not available and so the two concepts are different and are treated as such.

2.7 The Sensor Channel Transmittance

The sensor or measuring device that is used to determine the state of a dynamical system can be regarded as a noisy communication channel. Figure 2.1 is a possible representation for such a channel.



1. $\underline{S} = B(\underline{Y})$

2. $\underline{Z} = C(\underline{W})$

Figure 2.1. Interpretation of a Sensor as a Communication Channel.

The additive noise term, $\{n_k\}$ is usually an R^{th} order markoff process which is independent of the measured M^{th} order markoff process $\{y_k\}$. However, no markoffian properties will be assumed at this point. The signal shaping operator B is linear and physically realizable, and so it may be represented as a casual matrix having entries b_{ij}, such that $b_{ij} \equiv 0$, for all $i>j$. The only requirement

on this matrix operator is that DET [B] be non-zero. This is an obvious requirement that states mathematically the constraint, that it always be possible to completely recover the signal y, when the measurement noise is known.

The output filter C can be of a more general nature and, for convenience, it taken as a nonlinear casual operator of the form

$$\underline{Z} = C(\underline{W}) = \begin{bmatrix} c_1(w_1) \\ c_2(w_1,w_2) \\ c_3(w_1,w_2,w_3) \\ \cdot \\ \cdot \\ \cdot \\ c_K(w_1,\ldots,w_K) \end{bmatrix} \tag{2.7.1}$$

where

$$\underline{Z} = \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ z_K \end{bmatrix} \quad \text{and} \quad \underline{W} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \cdot \\ \cdot \\ \cdot \\ w_K \end{bmatrix} .$$

The "recovery" constraint on $C(\underline{W})$ requires that the Jacobian of the transformation, $J(\underline{W})$, must be non-zero for all $\underline{W}$. The Jacobian is given by the expression

$$J(\underline{W}) \triangleq \text{DET}\,[\Gamma(\underline{W})]$$

where $\Gamma(\underline{W})$ is the matrix of partial differentials given by

$$\Gamma(\underline{W}) = \frac{\partial c_i \; (w_1, \ldots, w_K)}{\partial w_j} = \{\gamma_{ij}\}.$$

In this dissertation the mutual information between the K dimensional input and output vectors is called "The Sensor Channel Transmittance." This quantity is given as:

$$I_K(\underline{Y};\underline{Z}) = H_K(\underline{Z}) - H_K(\underline{Z}/\underline{Y}) \tag{2.7.2}$$

or

$$I_K(\underline{Y};\underline{Z}) = \int p_z(\underline{Z}) \text{ LOG } \frac{1}{p_z(\underline{Z})} d\underline{Z} - \iint p(\underline{Z},\underline{Y}) \text{ LOG } \frac{1}{p(\underline{Z}/\underline{Y})} d\underline{Z}\, d\underline{Y} \tag{2.7.3}$$

where the integrals are of K^{th} order and are taken over the whole K dimensional space.

The probability density functions for $\underline{S}$ and $\underline{W}$ are related to the input and output density functions through the transformations:

$$\begin{aligned} &\text{i)} \quad \underline{Z} = C(\underline{W}) \\ &\text{ii)} \quad p_z(C(W)) = p_w(\underline{W})\,|J(\underline{W})|^{-1} \\ &\text{iii)} \quad p_y(\underline{Y}) = p_s(B^{-1}\underline{Y})\; B. \end{aligned} \tag{2.7.4}$$

The mutual information when written in terms of $\underline{S}$ and $\underline{W}$ becomes:

$$\begin{aligned} I_K(\underline{Y};\underline{Z}) = &\int p_w(\underline{W}) \text{ LOG } \frac{1}{p_w(\underline{W})} d\underline{W} + \int p_w(\underline{W}) \text{ LOG } |J(\underline{W})|\, d\underline{W} \\ &- \int p_{w,s}(\underline{W},\underline{S}) \text{ LOG } \frac{1}{p_{w,s}(\underline{W}/\underline{S})} d\underline{W}\, d\underline{S} \\ &- \iint p_{w,s}(\underline{W},\underline{S}) \text{ LOG } |J(\underline{W})|\, d\underline{W}\, d\underline{S} \end{aligned}$$

or

$$I_K(\underline{Y};\underline{Z}) = H_K(\underline{W}) - H_K(\underline{W}/\underline{S}). \tag{2.7.5}$$

The conditional entropy of $\underline{W}$ given $\underline{S}$ is simply $H_K(\underline{N})$. This may be shown as follows:

$$H_K(\underline{W}/\underline{S}) = \int_{-\infty}^{\infty} d\underline{S} \int_{-\infty}^{\infty} d\underline{W}\; p_{W,S}(\underline{W},\underline{S}) \text{ LOG } \frac{1}{p(\underline{W}/\underline{S})} \tag{2.7.6}$$

where

$$\underline{W} = \underline{S}+\underline{N}$$

$$p(\underline{W}/\underline{S}) = p_N(\underline{W}-\underline{S}) \tag{2.7.7}$$

$$p_{ws}(\underline{W},\underline{S}) = p_{ns}(\underline{W}-\underline{S},\underline{S}) = p_s(\underline{S})\; p_n(\underline{W}-\underline{S}) \tag{2.7.8}$$

and $p_n(\underline{N})$ is the probability density function of the K dimensional noise vector. Making the proper substitutions reduces the conditional entropy to

$$\begin{aligned} H_K(\underline{W}/\underline{S}) &= \int_{-\infty}^{\infty} d\underline{S} \int_{-\infty}^{\infty} d\underline{W}\; p_{ns}(\underline{W}-\underline{S},\underline{S}) \text{ LOG } \frac{1}{p_n(\underline{W}-\underline{S})} \\ &= \int_{-\infty}^{\infty} p_s(\underline{S})d\underline{S} \int_{-\infty}^{\infty} d\underline{N}\; p_n(\underline{N}) \text{ LOG } \frac{1}{p_n(\underline{N})} \\ &= H_K(\underline{N}) . \end{aligned} \tag{2.7.9}$$

Using this result in the expression for the mutual information yields:

$$I_K(\underline{Y};\underline{Z}) = H_K(\underline{S}+\underline{N}) - H_K(\underline{N}) . \tag{2.7.10}$$

Now using property 2 (Table I) of entropy, it is true that

$$H_K(C(\underline{N})) = H_K(\underline{N}) + \int_{-\infty}^{\infty} d\underline{N}\, p_n(\underline{N}) \,\mathrm{LOG}\, |J(\underline{N})|$$

In addition,

$$H_K(C(\underline{S}+\underline{N})) = H_K(\underline{S}+\underline{N}) + \int_{-\infty}^{\infty} d\underline{W}\, p_w(\underline{W}) \,\mathrm{LOG}\, |J(\underline{W})|$$

$$= H_K(\underline{S}+\underline{N}) + \int_{-\infty}^{\infty} d\underline{S} \int_{-\infty}^{\infty} d\underline{W}\, p_{ws}(\underline{W},\underline{S}) \,\mathrm{LOG}\, |J(\underline{W})|$$

$$= H_K(\underline{S}+\underline{N}) + \int_{-\infty}^{\infty} d\underline{S}\, p_s(\underline{S}) \int_{-\infty}^{\infty} d\underline{W}\, p_n(\underline{W}-\underline{S}) \,\mathrm{LOG}\, |J(\underline{W})|$$

$$= H_K(\underline{S}+\underline{N}) + \int_{-\infty}^{\infty} d\underline{N}\, p_n(\underline{N}) \,\mathrm{LOG}\, |J(\underline{N})| \,.$$

Then from these two equations and 2.7.10, it follows that

$$H_K(C(\underline{S}+\underline{N})) - H_K(C(\underline{N})) = H_K(\underline{S}+\underline{N}) - H_K(\underline{N})$$

$$H_K(C(\underline{S}+\underline{N})) - H_K(C(\underline{N})) = I_K(\underline{Y};\underline{Z}) \,. \qquad (2.7.11)$$

This form for the Channel Transmittance is more important than 2.7.10, because it relates directly to readily accessable quantities, i.e., the sensor output for no signal input (C(N)) and the sensor output for a signal input ($C(\underline{S}+\underline{N})$). Thus, according to equation (2.7.11), $I_K(\underline{Y};\underline{Z})$ can always be found for any sensor, even by simulation if necessary. So under laboratory conditions $I_K(\underline{Y};\underline{Z})$ may be determined experimentally, without any knowledge of B or C being required for this determination. This, coupled with the fact that all the bounding theorems on system performance subsequently proved

in the body of this dissertation, always depend directly in $I_K(\underline{Y};\underline{Z})$ [the term $H_K(\underline{N})$) never appears without the corresponding term $H_K(\underline{S}+\underline{N})$], means that it is never really necessary to know individually any of the sensor parameters [$\underline{N}$, $H_K(\underline{N})$, $H_K(\underline{S}+\underline{N})$, etc.] but only $I_K(\underline{Y};\underline{Z})$. Thus sensors which defy representation in conventional terms and which do not have suitable models can be completely described for the purposes of this discussion simply by a channel transmittance.

Recognizing that $C(\underline{S}+\underline{N})$ is the sensor output (or the signal measurements) it will not be surprising to learn below that the difference between the entropy of the output for signal and no-signal conditions

$$H_K(C(\underline{S}+\underline{N})) - H_K(C(\underline{N})) = \begin{array}{l}\text{The information in}\\ \text{the output about}\\ \text{the input}\end{array} \qquad (2.7.12)$$

can be interpreted as that amount of information in the output that is actually effective for reducing the entropy uncertainty of estimates of the input. Obviously if any estimates of a function of $\underline{Y}$ are to be made using the noisy measurement $\underline{Z}$, it is the Channel Transmittance $I_K(\underline{Y};\underline{Z})$ that will measure how successful the estimation procedure is. The reader should be cautioned that this is not a completely new interpretation of $I_K(\underline{Y};\underline{Z})$. If the sensor of Figure 2.1 was called a channel and if $\underline{Y}$ took on only a finite number of values and was called a code then there would be no objection to calling $I_K(\underline{Y};\underline{Z})$ the effectiveness of the decoding procedure. What is new is that this approach is now being applied to the estimation of continuous-state processes.

Since mutual information is always non-negative, equation 2.7.10 yields the important inequality

$$H_K(\underline{S}+\underline{N}) \geq H_K(\underline{N}) \tag{2.7.13}$$

In a similar manner, the mutual information $I_K(\underline{N};\underline{Z})$ yields an equally important inequality

$$I_K(\underline{N};\underline{Z}) = H_K(\underline{S}+\underline{N}) - H_K(\underline{S}) \geq 0 \tag{2.7.14}$$

or

$$H_K(\underline{S}+\underline{N}) \geq H_K(\underline{S}). \tag{2.7.15}$$

2.8 Entropy of the Sum Vector ($\underline{S}+\underline{N}$)

For many of the derivations to follow, it is necessary to know the conditional entropy of the sum of the two independently distributed random vectors $\underline{S}$ and $\underline{N}$, given all the past values of the sum. Since the probability densities of $\underline{S}$ and $\underline{N}$ are usually known in advance it is not difficult to conceive of using:

$$P(\underline{S}+\underline{N}) = \int_{-\infty}^{\infty} d\underline{\alpha}\; P_s(\underline{\alpha})\; P_n(\underline{S}+\underline{N}-\underline{\alpha}) \tag{2.8.1}$$

in order to determine the desired entropy function.

However it is sometimes desirable to be familiar with the properties of the conditional entropy without being obliged to carry out the convolution and entropy integrals directly. An expression similar to equation 2.5.6 is clearly desirable, but unfortunately, the sum of two markoff processes is no longer Markovian and no such simple equation

exists. However when the random variables involved are stationary and $\underline{N}$ is markoff then it is possible to determine a useful assymptotic description of the conditional entropy. It follows from the use of equation (2.3.5) that the mutual information increases monotonically with the number of samples used, i.e.,

$$
\begin{aligned}
I_K(\underline{S};\underline{S}+\underline{N}) &= I(s_1,s_2,\ldots,s_K;(s+n)_1,(s+n)2,\ldots,(s+n)_K) \\
&\leq I(s_1,s_2,\ldots,s_K,s_{K+1};(s+n)_1,(s+n)_2,\ldots,(s+n)_K) \\
&\leq I(s_1,s_2,\ldots,s_{K+1};(s+n)_1,\ldots,(s+n)_K,(s+n)_{K+1}) \\
&\leq I_{K+1}(\underline{S};\underline{S}+\underline{N}) \, . \qquad (2.8.2)
\end{aligned}
$$

Now expanding each of the mutual information terms according to a modification of equation (2.7.10),

$$I_K(\underline{S};\underline{S}+\underline{N}) = H_K(\underline{S}+\underline{N}) - H_K(\underline{N})$$

which leads to

$$H_K(\underline{S}+\underline{N}) - H_K(\underline{N}) \leq H_{K+1}(\underline{S}+\underline{N}) - H_{K+1}(\underline{N}) \, . \qquad (2.8.3)$$

It has already been determined (see Section 2.5, equation (2.5.3)) that the entropy of a K dimensional markoff noise vector is:

$$H_K(\underline{N}) = \sum_{k=1}^{K} H_1(\xi_k). \qquad (2.8.4)$$

This simplifies the above inequality to yield:

$$H_K(\underline{S}+\underline{N}) \leq H\left((s+n)_{K+1}/(s+n)_K(s+n)_{K-1},\ldots,(s+n)_1\right) + H_K(\underline{S}+\underline{N}) - H_1(\xi_{K+1}),$$

or finally

$$H\left|(s+n)_{K+1}/(\underline{S}+\underline{N})\right| \geq H_1(\xi_{K+1}).$$

In the stationary case this result, together with equation (2.2.19) leads to

$$H_1\left|(s+n)_{K+2}/(s+n)_{K+1},\ldots,(s+n)_2\right|$$

$$= H_1\left|(s+n)_{K+1}/(s+n)_K,\ldots,(s+n)_1\right|$$

$$\geq H_1\left|(s+n)_{K+2}/(s+n)_{K+1},\ldots,(s+n)_1\right| \geq H_1(\xi). \tag{2.8.5}$$

Thus the sequence $H\left|(s+n)_{K+1})/(\underline{S}+\underline{N})_K\right|$ is monotonically decreasing and is bounded below by $H_1(\xi)$. It is therefore convergent to a limit

$$\lim_{K\to\infty} H\left|(s+n)_{K+1}/(\underline{S}+\underline{N})\right| = H\left|(s_\infty+n_\infty)/(\underline{S}+\underline{N})\right| \geq H_1(\xi) \tag{2.8.6}$$

where in the limit $\underline{S}+\underline{N}$ becomes an infinite dimensional vector. An alternate and useful expression for the convergent property is:

Given any $\varepsilon>0$ there exists a κ sufficiently large so that for all $K\geq\kappa$,

$$H\left|(s+n)_{K+1}/(\underline{S}+\underline{N})\right| - H\left|(s_\infty+n_\infty)/(\underline{S}+\underline{N})\right| \leq \varepsilon. \tag{2.8.7}$$

This relationship will prove useful below for studying the incremental change in Channel Transmittance.

2.9 Incremental Channel Transmittance

According to equation (2.7.10) the K dimensional Sensor Channel Transmittance is

$$I_K(\underline{Y};\underline{Z}) = H_K(\underline{S}+\underline{N}) - H_K(\underline{N}). \tag{2.9.1}$$

As more data is taken and K increases, the Channel Transmittance (according to Theorem 2.3.1) must also increase and in general, it increases without bound. An interesting quantity to study is Δ_K, the Incremental Channel Transmittance, defined as

$$\Delta_K \triangleq I_{K+1}(\underline{Y};\underline{Z}) - I_K(\underline{Y};\underline{Z}) . \tag{2.9.2}$$

It follows from equation (2.8.2) that Δ_K is always positive and it represents the amount of "new" information obtained about <u>all</u> the signals (including the latest one) that may be derived from an additional measurement. After a sufficient length of time, the new data does not provide any new information about the oldest signal samples and therefore Δ_K approaches a steady-state limit Δ_∞. In the stationary case this statement is proven as follows.

From equations (2.9.1) and (2.9.2),

$$\Delta_K = H_{K+1}(\underline{S}+\underline{N}) - H_{K+1}(\underline{N}) - [H_K(\underline{S}+\underline{N}) - H_K(\underline{N})],$$

using equation (2.8.4) and

$$H_{K+1}(\underline{S}+\underline{N}) = H_K(\underline{S}+\underline{N}) + H_1\left((s+n)_{K+1}/(\underline{S}+\underline{N})_K\right) ,$$

the equation for Δ_K becomes

$$\Delta_K = H_1 \left((s+n)_{K+1}/(\underline{S}+\underline{N})_K\right) - H_1(\xi_{K+1}). \tag{2.9.3}$$

Then by applying equation (2.8.7) the convergent property of Δ_K may be stated as:

Given any $\varepsilon>0$ there exists a κ sufficiently large so that for all $K\geq\kappa$

$$\Delta_K - \Delta_\infty \leq \varepsilon \ . \qquad (2.9.4)$$

The asymptotically stationary Incremental Transmittance, Δ_∞, will prove useful in Section 3.8 for showing that in stationary markoff estimation problems the average entropy of the error approaches a steady-state value.

CHAPTER THREE
THE ESTIMATION PROBLEM

3.1 Summary

Except for the requirements of real time data processing, feedback control systems are very similar to estimating systems. Therefore it is not illogical to assume that the investigation of feedback should begin with an investigation of estimation, especially since this type of problem has received much attention and several mean square error solutions are known to exist [34, 35].

Useful entropy analysis is not just a matter of writing the system equations, transforming the corresponding probability density functions and then determining the various signal entropies. The scalar feedback problem studied in 3.2 best demonstrates the difficulties involved in this approach. The major result, that of bounding the error entropy from below is

$$H(x) \geq H(\frac{1}{1+a}\, y).$$

If "a" were known, then this bound could be determined; but then if "a" were known H(x) could be determined exactly. Thus in this example no useful purpose has been served by using the entropy measure. This setback is still further incentive for first developing entropy techniques by solving the simpler problem of estimating before attempting to solve the very complex feedback control problem.

The critical step for initiating entropy analysis, the study of the mutual information between the system error and the sensor output,

is investigated in 3.4 (after a suitable problem definition is given in 3.3). The most useful result of the mutual information approach is that it leads to a bound on the estimation error entropy that is independent of the estimating filter and is a function solely of the known properties of the system input and the system sensor.

In 3.5 a corollary is proven that is a reinterpretation of the estimation theorem through the use of the concept of a Sensor Channel Transmittance. This corollary completely relaxes all of the constraints on the form that the estimating system may take on. The filters "C" and "D" need not preserve information and the only description of the sensor that is now required for an entropy analysis is its Channel Transmittance, $I(\underline{Y};\underline{Z})$.

A simple example, presented in 3.6, is used to show that in the Gaussian-linear-estimation problem, entropy solutions lead to conventional results. With this experience as background, another corollary to the estimation theorem is proven, the results of which demonstrate clearly the basic similarities between variance and entropy as uncertainty measures. For linear Gaussian problems variance and entropy are completely interchangeable. The remainder of the chapter is devoted to understanding the implications of the results as they apply to the change in error entropy with additional measurements.

3.2 <u>Naive Application of Entropy Analysis</u>

The tracking problem produces an interesting example of how undisciplined use of the entropy criterion can lead to difficulties.

Consider the scalar feedback estimation problem shown in Figure 3.1. The following theorem applies:

Theorem 3.2

The entropy of x is bounded by:

$$H(x) \geq H\left(\frac{1}{a+1} n\right) \qquad 3.2.1$$

i.e., the entropy of x always exceeds the entropy of the component of the noise that appears at the output.

Proof:

Since

$$x = \frac{1}{1+a} y - \frac{a}{1+a} n$$

$$y = y,$$

it follows that the transformation of the probability density functions is

$$p_{xy}(x,y) = p_{ny}(n,y) \left(\frac{1}{1+a}\right)^{-1}.$$

So that using property 2 (Table I), the joint entropy of x and y may be found in terms of the entropy of y and n, i.e.,

$$H(x,y) = H(n,y) + \mathrm{LOG} \frac{a}{1+a}$$

$$= H(y) + H(n) + \mathrm{LOG} \frac{a}{1+a}$$

$$= H(y) + H\left(\frac{a}{1+a} n\right).$$

Now examine the mutual information between x and y:

$$I(x;y) = H(x) + H(y) - H(x,y)$$

$$= H(x) + H(y) - H(y) - H\left(\frac{a}{1+a} n\right),$$

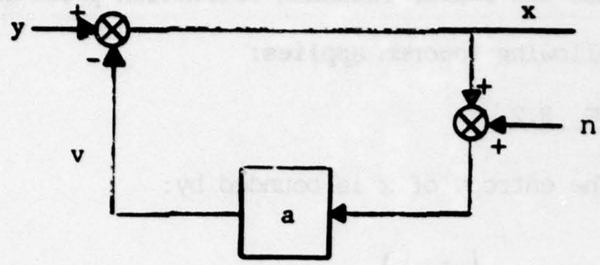


Figure 3.1. A Scalar Feedback Problem.

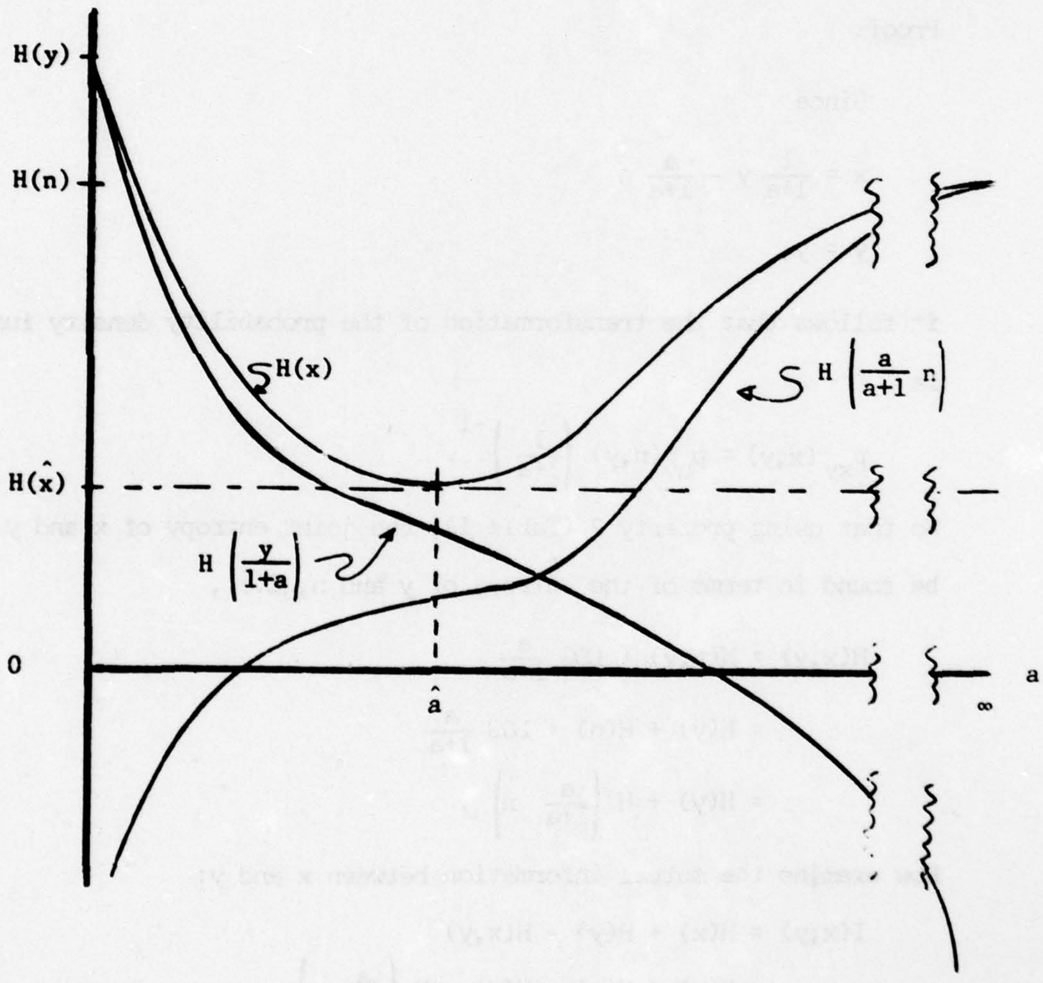


Figure 3.2. A Typical Graph of H(x) vs. a.

or

$$H(x) = I(x;y) + H\left(\frac{a}{1+a}\,n\right).$$

Since I(x;y) is always positive, the theorem is proven.

In an analogous manner, using I(x;n), it is possible to prove that the error entropy always exceeds the entropy of the closed loop component of the signal at the output, i.e.,

$$H(x) \geq H\left(\frac{1}{1+a}\,y\right).$$

It would appear that these bounds on error entropy would have strong applications for system analysis. Unfortunately they do not. The reason is that even if the minimum possible error entropy occurs with satisfaction of the equality in equation 3.2.1 (and it definitely does not), evaluation of the quantity $H\left(\frac{a}{1+a}\,n\right)$ still depends on determining the values of "a" to achieve the minimum, and if "a" were available, entropy analysis would have no advantage over the actual calculation of the properties of the true error.

In actuality $H\left(\frac{a}{1+a}\,n\right)$ $\left[\text{or } H\left(\frac{1}{1+a}\,y\right)\right]$ have little bearing on the true minimum error entropy that may be achieved by this system. A typical relationship between "a" and error entropy is shown in Figure 3.2. Thus, there is no simple way to find either $\hat{a}$ or $H(\hat{x})$ from this particular application of entropy for feedback analysis, even though two apparently very nice reuslts were obtained in the process.

3.3 The Estimation Problem; Description

As the first useful step in the study of the use of information theory, it is instructive to investigate the elementary estimation

problem shown in Figure 3.3. A random signal sequence taking on the values y_k, k=1,2,...,K is processed by the known dynamical system D. D may be nonlinear and time varying and it may also be non-realizable in the sense that the system output could depend on future values of the input. A suitable vector representation for this operation can be written as:

$$\underline{U} = D(\underline{Y}),$$

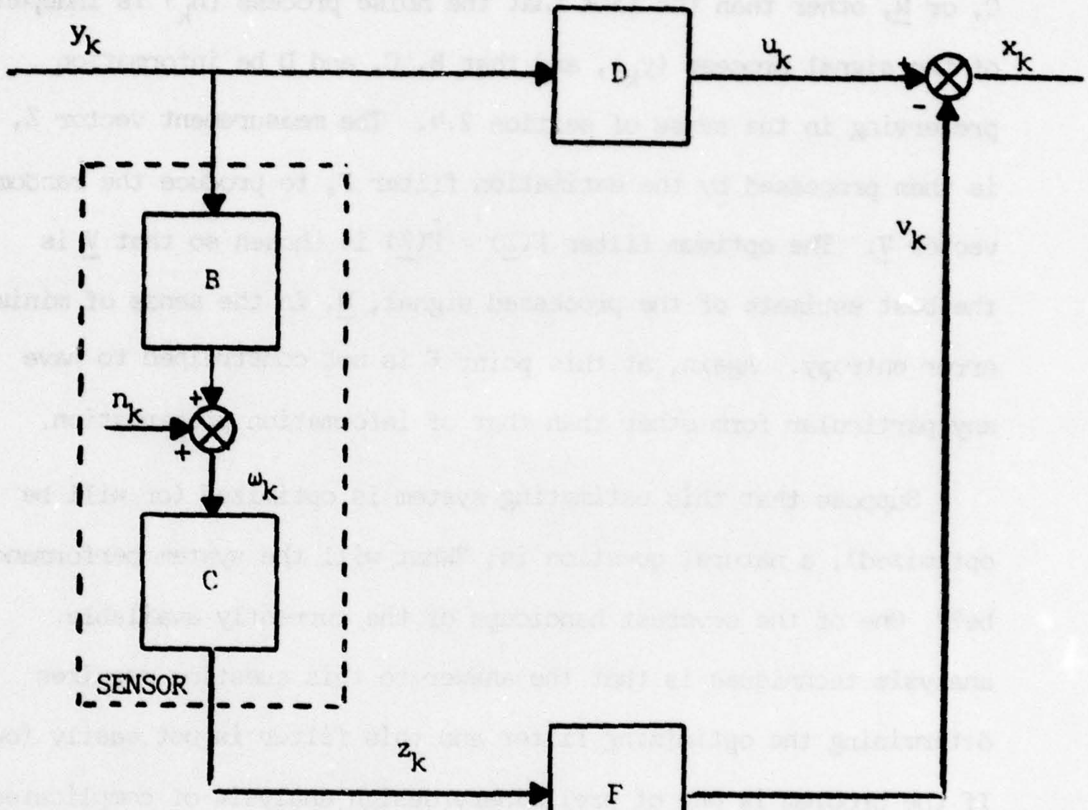
where the $\underline{U}$ and $\underline{Y}$ vectors are K dimensional vectors whose components are the members of the corresponding random sequence, i.e.,

$$\underline{U} = \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_k \\ \cdot \\ \cdot \\ \cdot \\ u_K \end{bmatrix} \qquad \underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_k \\ \cdot \\ \cdot \\ \cdot \\ y_K \end{bmatrix}$$

and

$$D(\underline{Y}) = \begin{bmatrix} d_1(y_1,y_2,\ldots,y_K) \\ d_2(y_1,y_2,\ldots,y_K) \\ \cdot \\ \cdot \\ \cdot \\ d_K(y_1,y_2,\ldots,y_K) \end{bmatrix} = \begin{bmatrix} d_1(\underline{Y}) \\ d_2(\underline{Y}) \\ \cdot \\ \cdot \\ \cdot \\ d_K(\underline{Y}) \end{bmatrix}$$

if $d_k(y_1,y_2,\ldots,y_k,\ldots,y_K)$ is independent of all y_i for $i \geq k+1$, then the system is called casual.



1. $\underline{U} = D(\underline{Y})$
2. $\underline{Z} = C(B(\underline{Y}) + \underline{N})$
3. $\underline{V} = F(\underline{Z})$
4. $\underline{X} = \underline{U} - \underline{V}$

Figure 3.3. The Estimation Problem.

At the same time, the sequence $\{y_k\}$ is measured by the noisy sensor consisting of a nonlinear pre-filter, B, the corrupting additive noise $\underline{N}$ and the nonlinear post-filter C. At this point in the development, no other constraints are placed on the forms of B, C, or $\underline{N}$, other than the fact that the noise process $\{n_k\}$ is independent of the signal process $\{y_k\}$, and that B, C, and D be information preserving in the sense of section 2.4. The measurement vector Z, is then processed by the estimation filter F, to produce the random vector $\underline{V}$. The optimum filter $F(\underline{Z}) = \hat{F}(\underline{Z})$ is chosen so that $\underline{V}$ is the best estimate of the processed signal, $\underline{U}$, in the sense of minimum error entropy. Again, at this point F is not constrained to have any particular form other than that of information preservation.

Suppose that this estimating system is optimized (or will be optimized), a natural question is; "What will the system performance be?" One of the severest handicaps of the currently available analysis techniques is that the answer to this question requires determining the optimizing filter and this filter is not easily found. If the problem is one of preliminary design analysis of complicated systems, then not being able to easily calculate the optimum performance adds greatly to the design problem and obscures the importance of certain system parameters. An advantage of the information theory approach developed herein is that it provides the "back door" to system performance evaluation without solving the filter problem directly. When the estimation problem has the form shown in Figure 3.3, the following theorem is applicable.

3.4 The Entropy Theorem for Estimation

Theorem 3.4:

1. For the general estimation problem shown in Figure 3.3, and an arbitrary filter function, $F(\underline{Z})$, if all the transformations are information preserving, then the entropy of the error vector $\underline{X}$ always satisfies the inequality:

$$H(\underline{X}) \geq H(D(\underline{Y})) + H(N) - H(B(\underline{Y}) + \underline{N}) = H_o \, , \tag{3.4.1}$$

where H_o is independent of $F(\underline{Z})$. In other words, the reduction in the processed signal entropy, $H(D(\underline{Y})) - H(\underline{X})$, due to feed-forward estimation, cannot exceed the Sensor Channel Transmittance, $I(\underline{Y};\underline{Z})$, where

$$I(\underline{Y};\underline{W}) = I(\underline{Y};\underline{Z}) = H(B(\underline{Y}) + \underline{N}) - H(\underline{N}).$$

2. Minimizing the mutual information $I(\underline{X};\underline{W})$ is equivalent to minimizing the error vector entropy.

3. The minimum error entropy occurs when $I(\underline{X};\underline{W}) \equiv 0$. This minimum entropy is

$$H(\underline{X})\Big]_{min} = H_o = H(D(\underline{Y})) + H(\underline{N}) - H(B(\underline{Y}) + \underline{N}) \tag{3.4.2}$$

and is attainable if the optimum filter $\hat{F}$ can be chosen so that the random vectors $\underline{X}$ and $\underline{W}$ are independent.

The proof of this theorem is based on considering the mutual information between the measurement vector, $\underline{W}$, and the error vector $\underline{X}$. It is intuitively obvious that in the sense of entropy, if the system is optimized [$H(\underline{X})$ is minimum] then $\underline{V}$ can not contain any

irreducible information about the error $\underline{X}$, i.e., $I(\underline{X};\underline{V})$ must be a minimum. Since $I(\underline{X};\underline{V}) = I(\underline{X};\underline{W})$ for information preservation, it follows that this implies that the mutual information between the error and the measurements is a minimum. If it is not a minimum then $\hat{\underline{Z}}$ could be reprocessed and this additional information removed, thus reducing the entropy of $\underline{X}$ further. This intuitive notion that the minimum error entropy corresponds to minimum "error-measurements" mutual information will be proven.

Proof:

The mutual information between $\underline{X}$ and $\underline{W}$ is defined as:

$$I(\underline{X};\underline{W}) = \int d\underline{X}\ p_{xw}(\underline{X},\underline{W}) \ \mathrm{LOG} \frac{p_{xw}(\underline{X},\underline{W})}{p_x(\underline{X})p_w(\underline{W})} , \qquad (3.4.3)$$

where $p_{xw}(\underline{X},\underline{W})$ is a function of 2K arguments and each integral is K dimensional, i.e., $d\underline{X} = dx_1\ dx_2 \ldots dx_K$. According to Table I, Property 15, this mutual information may be written in terms of the individual entropies of $\underline{X}$ and $\underline{W}$ as

$$I(\underline{X};\underline{W}) = H(\underline{X}) - H(\underline{X},\underline{W}) + H(\underline{W}) . \qquad (3.4.4)$$

The vector $\underline{W}$ is given by

$$\underline{W} = \underline{N} + B(\underline{Y}),$$

so it follows that

$$H(\underline{W}) = H(\underline{N} + B(\underline{Y})).$$

If the distributions of $\underline{N}$ and $\underline{Y}$ are known, then at least in theory $H(\underline{W})$ may always be calculated.

The joint entropy between $\underline{X}$ and $\underline{W}$ is found from examination of the system equations

$$\underline{W} = \underline{N} + B(\underline{Y}) \tag{3.4.5a}$$

$$\underline{X} = -F(C(\underline{N}+B(\underline{Y}))) + D(\underline{Y}) = -F_c(\underline{N}+B(\underline{Y})) + D(\underline{Y}) \tag{3.4.5b}$$

The Jacobian J, for this transformation of $\begin{bmatrix} \underline{N} \\ \underline{Y} \end{bmatrix}$ into $\begin{bmatrix} \underline{W} \\ \underline{X} \end{bmatrix}$ is given by the expression

$$J \triangleq J(\underline{N};\underline{Y}) = \mathrm{DET}\left[\begin{array}{c|c} \partial_n N & +\partial_y B(\underline{Y}) \\ \hline -\partial_n F_c & \partial_y D - \partial_y F_c \end{array}\right]$$

$$= \mathrm{DET}\left[\begin{array}{c|c} I & +\partial_y B(\underline{Y}) \\ \hline -\partial_n F_c & \partial_y D - \partial_y F_c \end{array}\right]$$

$$= \mathrm{DET}\,[\partial_y D - \partial_y F_c + (\partial_n F_c)(\partial_y B(\underline{Y}))] \tag{3.4.6}$$

where the entries in the partitioned matrix are defined in the following manner:

$\partial_n N = I$ (the identity matrix)

$\partial_y B(\underline{Y}) = B'$ (with $B'_{ij} = \dfrac{\partial b_i(\underline{Y})}{\partial y_i}$) .

$$\partial_n F_c \triangleq \begin{bmatrix} \dfrac{\partial F_{c_1}}{\partial n_1}(\underline{N}+B(\underline{Y})) & ,\ldots, & \dfrac{\partial F_{c_1}}{\partial n_K}(\underline{N}+B(\underline{Y})) \\ \vdots & & \vdots \\ \dfrac{\partial F_{c_K}}{\partial n_1}(\underline{N}+B(\underline{Y})) & ,\ldots, & \dfrac{\partial F_{c_K}}{\partial n_K}(\underline{N}+B(\underline{Y})) \end{bmatrix}$$

$$\partial_y F_c \triangleq \begin{bmatrix} \dfrac{\partial F_{c_1}(\underline{N}+B(\underline{Y}))}{\partial y_1}, & \ldots, & \dfrac{\partial F_{c_1}(\underline{N}+B(\underline{Y}))}{\partial y_K} \\ \vdots & & \vdots \\ \dfrac{\partial F_{c_k}(\underline{N}+B(\underline{Y}))}{\partial y_1}, & \ldots, & \dfrac{\partial F_{c_K}(\underline{N}+B(\underline{Y}))}{\partial y_K} \end{bmatrix} = \partial_n F_c B'$$

$$\partial_y D \triangleq \begin{bmatrix} \dfrac{\partial d_1}{\partial y_1}, & \dfrac{\partial d_1}{\partial y_2} & \ldots, & \dfrac{\partial d_1}{\partial y_K} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial d_K}{\partial y_1}, & \dfrac{\partial d_K}{\partial y_Z} & \ldots, & \dfrac{\partial d_K}{\partial y_K} \end{bmatrix}$$

Using this notation,

$$J(\underline{N},\underline{Y}) = \text{DET}\,[\partial_n F_c]\ \text{DET}\,[B'] + \text{DET}\,[\partial_y D] - \text{DET}\,[\partial_y F_c],$$

but it is obvious that

$$\text{DET}[\partial_y F_c] = \text{DET}\,[(\partial_n F_c)\,B'] = \text{DET}\,[\partial_n F_c]\ \text{DET}\,B'],$$

so that finally,

$$J(\underline{N},\underline{Y}) = \text{DET}\,[\partial_y D(\underline{Y})] \qquad (3.4.7)$$

The Jacobian given in this manner may then be used to express the probability relationship existing between the vectors $(\underline{N},\underline{Y})$ and $(\underline{W},\underline{X})$.

The pertinent probability density function equations are:

$$p_{xw}(\underline{X},\underline{W})\ d\underline{X}\ d\underline{W} = p_{ny}\ (\underline{N},\underline{Y})\ d\underline{Y}\ d\underline{N} = p_y(\underline{Y})\ p_n(\underline{N})\ d\underline{Y}\ d\underline{N} \quad (3.4.8)$$

and

$$p_{xw}(\underline{X},\underline{W}) = p_y(\underline{Y}(\underline{X},\underline{W}))\ p_n(\underline{N}(\underline{X},\underline{W}))\ |J(\underline{N},\underline{Y})|^{-1}. \quad (3.4.9)$$

The joint entropy of $(\underline{X},\underline{W})$ is defined as

$$H(\underline{X},\underline{W}) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\underline{X}\ d\underline{W}\ p_{xw}(\underline{X},\underline{W})\ \mathrm{LOG}\ \frac{1}{p_{xw}(\underline{X},\underline{W})}\ .$$

When 3.4.8 and 3.4.9 are substituted into this expression it becomes

$$H(\underline{X},\underline{W}) = \int_{\infty}^{\infty}\int_{\infty}^{\infty} d\underline{N}\ d\underline{Y}\ \ p_n(\underline{N})\ p_y(\underline{Y})\ \mathrm{LOG}\ \frac{1}{p_n(\underline{N})\ p_y(\underline{Y})\ |\mathrm{DET}[\partial_y D]\ |^{-1}}$$

(3.4.10)

If the log term is expanded and the indicated integrations performed, the mutual entropy can then be written as:

$$H(\underline{X},\underline{W}) = H(\underline{Y}) + E_y\ \{\mathrm{LOG}\ |\mathrm{DET}\ [\partial_y D]|\} + H(\underline{N}) \quad (3.4.11)$$

where E_y { } is the expectation operation taken with respect to the random variable y, i.e.,

$$E_y\ \{\underline{\alpha}\} \triangleq \int_{\infty}^{\infty} d\underline{Y}\ \underline{\alpha}\ p_y(\underline{Y})\ . \quad (3.4.12)$$

An important simplification of this equation is possible, if it is noted that

$$H(D(\underline{Y})) = H(\underline{Y}) + E_y\ \{\mathrm{LOG\ DET}\ [\partial_y D]\}. \quad (3.4.13)$$

With the aid of equation (3.4.13) the joint entropy of $\underline{X}$ and $\underline{W}$ can be expressed in terms of the entropies of $\underline{Y}$ and $\underline{N}$,

$$H(\underline{X},\underline{W}) = H(D(\underline{Y})) + H(\underline{N}). \tag{3.4.14}$$

The expressions for $H(\underline{W})$ and $H(\underline{X},\underline{W})$ may now be combined in equation 3.4.4 to yield the equation from which all the results of this theorem follow:

$$I(\underline{X};\underline{W}) = H(\underline{X}) - H(D(\underline{Y})) - H(\underline{N}) + H(B(\underline{Y})+\underline{N}) . \tag{3.4.15}$$

Only $I(\underline{X};\underline{W})$ and $H(\underline{X})$ are functions of the estimation filter $F(\underline{Z})$, therefore, minimum error entropy occurs for minimum mutual information. Since mutual information is always a non-negative quantity it must be true that

$$H(\underline{X}) \geq H(D(\underline{Y})) + H(\underline{N}) - H(B(\underline{Y})+\underline{N}) . \tag{3.4.16}$$

The importance of this equation is that it relates the entropy of $\underline{X}$ to the known entropies of the input quantities, without requiring explicit knowledge of the estimating filter $F(\underline{Z})$. In addition the relationship is true for all possible filters. Let

$$H(D(\underline{Y})) + H(\underline{N}) - H(B(\underline{Y})+\underline{N}) = H_o. \tag{3.4.17}$$

H_o is a constant for a given sensor and a given dynamical system D and is not a function of $F(\underline{Z})$. Then from equation 3.4.16, it follows that:

$$H(\underline{X}) \geq H_o. \tag{3.4.18}$$

$H(\underline{X})$ is bounded below and can never be less than H_o.

If the equality in equation 3.4.18 can be achieved for some choice of $F(\underline{Z})$ then that filter must be the optimum entropy filter because is causes $H(\underline{X})$ to take on its minimum value. The equality of equation 3.4.16 can be attained only if $I(\underline{X};\underline{W}) = 0$, but this occurs if and only if $\underline{X}$ and $\underline{W}$ are independent. Therefore the optimum estimating filter, $\hat{F}(\underline{Z})$ is such that

$$p_{XW}(\underline{X},\underline{W}) = p_X(\underline{X})\, p_W(\underline{W}) \tag{3.4.19}$$

with the result that

$$\min_{F(\underline{Z})} \{H(\underline{X})\} \triangleq H(\hat{\underline{X}}) = H_o \quad . \tag{3.4.20}$$

It is impossible to over emphasize the fact that in either case, that of determining the entropy lower bound for estimation with sub-optimal filters, or of determining the minimum entropy for optimum estimation, it is not required to actually know either $F(\underline{Z})$, or $\hat{F}(\underline{Z})$.

The requirement that $\underline{X}$ be independent of $\underline{W}$ requires that all the data be made available before an estimation is made. This is analogous to a similar requirement of the Shannon coding theorem, and it follows that if minimum errors are to be achieved for either optimum estimation or optimum coding it is required that the entire message (signal) be known before it is coded (filtered).

3.5 The Channel Transmittance Approach to Estimation

Having shown that the performance of an estimating system that utilizes an additive noise sensor is bounded by the Channel

Transmittance of that sensor, it is easy to conjecture that only the informational properties of the sensor are important for entropy analysis and the particular form taken by the sensor's model is not significant. This relaxation of the constraints of Theorem 3.4 is stated rigorously in the following theorem.

Theorem 3.5 (A stronger version of Theorem 3.4)

If a signal vector $\underline{Y}$, is measured by a sensor having a Channel Transmittance, $I(\underline{Y};\underline{Z})$ then the entropy of the error in estimating $D(\underline{Y})$ is $H(\underline{X})$, and $H(\underline{X})$ is bounded as

$$H(\underline{X}) \geq - I(\underline{Y};\underline{Z}) + I(\underline{X};\underline{Z}) + H(D(\underline{Y})). \tag{3.5.1}$$

Both $F(\underline{Z})$ and $D(\underline{Y})$ are completely arbitrary single value functions, and the equality holds when $D(\underline{Y})$ preserves information. When $F(\underline{Z})$ is chosen so that

$$I(\underline{X};\underline{Z}) = 0 \tag{3.5.2}$$

this corresponds to the minimum error entropy,

$$\underset{F(\underline{Z})}{\text{MIN}} \; \{H(\underline{X})\} = \hat{H}(\underline{X}) = H(D(\underline{Y})) - I(\underline{Y};\underline{Z}). \tag{3.5.3}$$

Thus the maximum system performance improvement $H(D(\underline{Y})) - H(\underline{X})$ is equal to the Sensor Channel Transmittance. Note that this theorem relaxes the constraints on the estimation problem to such an extent that now all the components may have arbitrary form.

Proof:

Begin with the system equations defined by Figure 3.4.

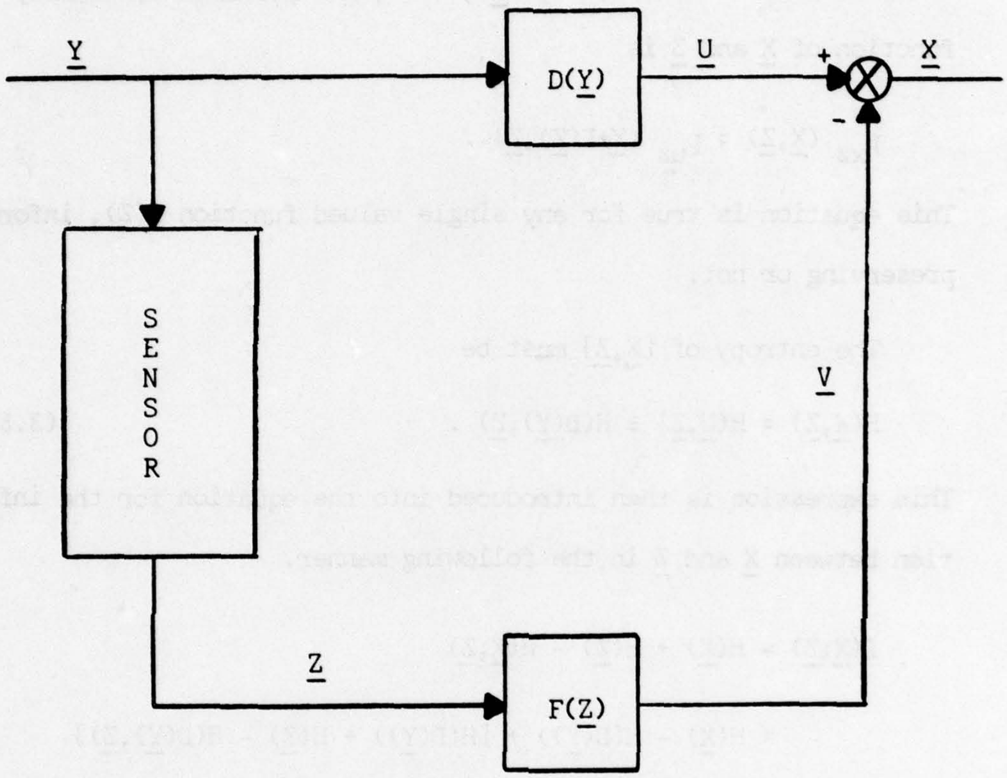


Figure 3.4. The Estimation Problem with Generalized Sensor.

$$\begin{bmatrix} \underline{X} = D(\underline{Y}) - F(\underline{Z}) \\ \underline{Z} = 0 + Z \end{bmatrix} \tag{3.5.4}$$

It then follows that with $U=D(\underline{Y})$, the joint probability density function of $\underline{X}$ and $\underline{Z}$ is

$$p_{xz}(\underline{X},\underline{Z}) = p_{uz}(\underline{X}+F(\underline{Z}),\underline{Z}) .$$

This equation is true for any single valued function $F(\underline{Z})$, information preserving or not.

The entropy of $[\underline{X},\underline{Z}]$ must be

$$H(\underline{X},\underline{Z}) = H(\underline{U},\underline{Z}) \equiv H(D(\underline{Y}),\underline{Z}) . \tag{3.5.6}$$

This expression is then introduced into the equation for the information between $\underline{X}$ and $\underline{Z}$ in the following manner.

$$\begin{aligned} I(\underline{X};\underline{Z}) &= H(\underline{X}) + H(\underline{Z}) - H(\underline{X},\underline{Z}) \\ &= H(\underline{X}) - H(D(\underline{Y})) + [H(D(\underline{Y})) + H(\underline{Z}) - H(D(\underline{Y}),\underline{Z})] \\ &= H(\underline{X}) - H(D(\underline{Y})) + I(D(\underline{Y});\underline{Z}) \end{aligned} \tag{3.5.7}$$

but

$$I(\underline{Y};\underline{Z}) \geq I(D(\underline{Y});\underline{Z}) \tag{3.5.8}$$

with the equality holding when the transformation $D(\underline{Y})$ preserves information. Therefore

$$H(\underline{X}) \geq H(D(\underline{Y})) + I(\underline{X};\underline{Z}) - I(\underline{Y};\underline{Z}) \tag{3.5.9}$$

and the theorem is proven.

3.6 Example

An example supporting the major conjectures of this theorem is had by considering the simple (and classical) problem of estimation of a Gaussian signal in additive Gaussian noise. Then:

$$D = I$$

$$B = I$$

$$C = I$$

$$\underline{Z} = \underline{Y} + \underline{N}$$

and $\hat{F}(\underline{Z})$ will obviously be a linear function of $\underline{Z}$.

The optimum mean square filter must satisfy the requirement that the minimum error be orthogonal to be signal $\underline{Y} + \underline{N}$ [42, p.218]. This leads to the equations

$$E\,\{(\underline{Y} - \hat{F}\,\underline{Z})\ \ \underline{Z}^T\} = 0$$

$$(\hat{F})\ E\,\{\underline{ZZ}^T\} = E\,\{\underline{YZ}^T\}\ .$$

For convenience all variables are taken with zero means. Using the notation

$$R_y \triangleq E\,\{\underline{YY}^T\}$$

$$R_n \triangleq E\,\{\underline{NN}^T\}$$

the optimum filter is found to be the matrix

$$\hat{F} = R_y\,[R_y+R_n]^{-1}\ .$$

The minimum mean square error matrix is defined to be

$$E\{\underline{\hat{X}}\underline{\hat{X}}^T\} \triangleq \sigma^2 = E\ (\underline{Y}-\hat{F}\underline{Z})(\underline{Y}-\hat{F}\underline{Z})^T = R_y - \hat{F}R_y$$

$$= R_y - R_y\,[R_y+R_n]^{-1} R_y$$

$$= R_y(1 - [R_y+R_n]^{-1} R_y)$$

Simplification of this last expression leads to

$$\sigma^2 = R_y\ [R_y+R_n]^{-1}\ (R_y+R_n-R_y)$$

$$\sigma^2 = R_y\ [R_y+R_n]^{-1}\ R_n$$

Since $\underline{\hat{X}}$, the optimum error vector is Gaussian, the error entropy may be written directly as

$$H(\underline{\hat{X}}) = \frac{1}{2} \text{LOG } \{(2\pi)^K \text{ DET } [\sigma^2]\}$$

or, making use of σ^2 found above, $H(\underline{\hat{X}})$ can be expanded to yield

$$H(\underline{\hat{X}}) = \frac{1}{2} \text{LOG } \{(2\pi)^K \text{ DET } [R_y]\}$$

$$+ \frac{1}{2} \text{LOG } \{(2\pi)^K \text{ DET } [R_y+R_n]^{-1}\}$$

$$+ \frac{1}{2} \text{LOG } \{(2\pi)^K \text{ DET } [R_n]\}.$$

Recognizing each of the terms on the right side of this equation leads to

$$H(\underline{\hat{X}}) = H(\underline{Y}) + H(\underline{N}) - H(\underline{Y}+\underline{N}),$$

which is, of course, the major conclusion of the theorem.

Because of the unique relationship between Gaussian variance and Gaussian entropy it is suspected that an entropy expression exists which unites the two points of view. The following corollary to Theorem 3.5 provides this unification by considering the signal conditioned by the measurements as the fundamental analysis quantity.

3.7 Corollary I to the Estimation Theorem 3.5

Corollary I:

The entropy of the error vector always exceeds the entropy of the processed signal conditioned on the measurements,

$$H(\underline{X}) \geq H(D(\underline{Y})/\underline{Z}) \qquad (3.7.1)$$

with equality if the optimum filter is used.

Proof:

From equation (3.5.7) the error entropy is

$$H(\underline{X}) = H(D(\underline{Y})) + I(\underline{X},\underline{Z}) - I(D(\underline{Y}),\underline{Z})$$

using

$$I(D(\underline{Y}),\underline{Z}) = H(D(\underline{Y})) - H(D(\underline{Y})/\underline{Z})$$

and

$$I(\underline{X},\underline{Z}) \geq 0.$$

It follows that

$$H(\underline{X}) \geq H(D(\underline{Y})/\underline{Z})$$

where the equality holds when $F(\underline{Z})$ is chosen so that

$$I(\underline{X},\underline{Z}) = 0,$$

thus proving the corollary.

This result resembles very much the accepted relationships of classical Gaussian mean square analysis. In fact, in the special case of Gaussian random variables in a linear system the corollary reduces to

$$\mathrm{DET}\,[\mathrm{VAR}\,\{\underline{X}\}] = \mathrm{DET}\,[\mathrm{VAR}\,\{D(\underline{Y})/\underline{Z}\}] \tag{3.7.2}$$

an equation which is well known [45, p.225]. Of course this is just a practical example of the supposition first made by DeGroot [20] to the effect that in uncertainty analysis the various measures, such as variance and entropy, are interchangeable.

3.8 Steady-State Entropy

When the signal processes are stationary and the transformations are time invariant, it must follow that after a sufficiently long time steady-state conditions will exist. Under such conditions it would be useful to know the entropy of one coordinate of the error vector. This knowledge is not directly available since all the previous entropy expressions are always written for vector quantities. However the average entropy per coordinate $\frac{1}{K} H_K(\underline{X})$, will serve the same purpose.

Begin with equation (3.4.1), using $\underline{S} = B(\underline{Y})$

$$\frac{1}{K} H_K(\underline{X}) \geq \frac{1}{K} H_K(D(\underline{Y})) + \frac{1}{K} [H_K(\underline{N}) - H_K(\underline{S}+\underline{N})] \tag{3.8.1}$$

by assumption

$$\lim_{K\to\infty} \frac{1}{K} H_K(D(\underline{Y}))$$

must exist and it will have the value H(u), i.e., it is the entropy of the steady-state processed signal.

According to the definition of Incremental Channel Transmittance given in Chapter Two, it is possible to write

$$H_K(\underline{N}) - H_K(\underline{S}+\underline{N}) = -\left[I_M(\underline{Y};\underline{Z}) + \sum_{M+1}^{K} \Delta_i \right] . \tag{3.8.2}$$

According to the result presented in equation (2.9.4), given any ε, arbitrarily small, it is always possible to find κ sufficiently large to ensure that

$$\varepsilon \geq \varepsilon_i \qquad i > \kappa$$

where ε_i is defined as

$$\varepsilon_i \triangleq \Delta_i - \Delta_\infty \; . \tag{3.8.3}$$

M is now fixed so that

$$M > \kappa \; .$$

Combining equations (3.8.1), (3.8.2), (3.8.3), and taking the limit yields

$$\lim_{K\to\infty} \frac{H(\underline{X})}{K} \geq H(u) - \lim_{K\to\infty} \frac{I_M(\underline{Y};\underline{Z})}{K} - \lim_{K\to\infty} \frac{K-M-1}{K} \Delta_\infty \lim_{K\to\infty} \sum_{M+1}^{K} \frac{\varepsilon_i}{K}$$

$$\lim_{K\to\infty} \frac{H_K(\underline{X})}{K} \geq H(u) - \Delta_\infty - \lim_{K\to\infty} \frac{K-M-1}{K} \varepsilon$$

or

$$\left[H(u) - \lim_{K} \frac{H_K(\underline{X})}{K} \right] - \Delta_\infty \leq \varepsilon$$

which is the requirement that the steady-state system performance improvement due to feed-forward estimation be limited by the asymptotically stationary Incremental Channel Transmittance, Δ_∞.

CHAPTER FOUR

THE FEEDBACK CONTROL PROBLEM

4.1 Summary

The most important application of the techniques of entropy analysis is in regard to feedback control systems. In this chapter and the next, two types of closed loop control systems will be investigated and a completely new interpretation of feedback control will be derived.

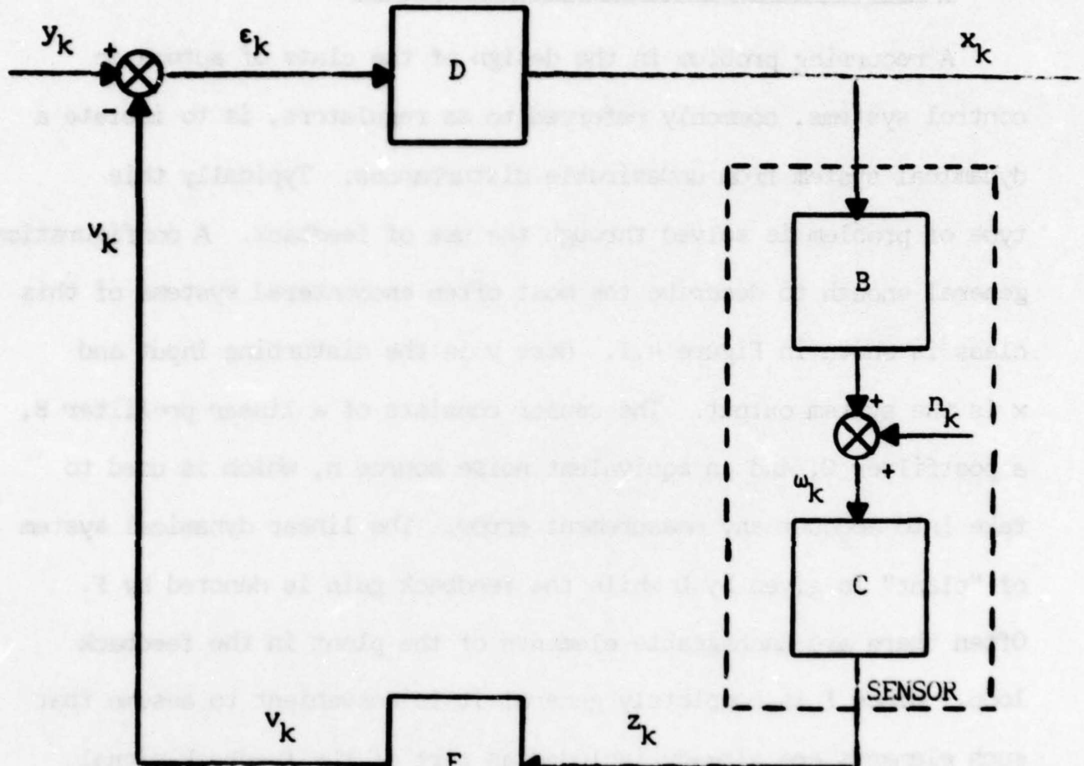
This present chapter is concerned with the performance of the disturbance rejecting system shown in Figure 4.1 which maintains the minimum entropy output, despite the presence of both external noise y and sensor noise n. The major result of this chapter is theorem 4.4 which can be stated mathematically as

$$H(D\underline{Y}) - H(\underline{X}) = I(\underline{U};\underline{Z}) - I(\underline{X};\underline{Z}) .$$

This equation leads to an absolute description of the entropy of the error vector $\underline{X}$ and to a bound on the improvement of system performance, due the use of feedback, that is the channel property of the system sensor. Also from this equation follows the corollary:

$$H(\underline{X}) \geq H(D\underline{Y}/\underline{Z}).$$

Evaluation of the resulting equations with regard to actually achieving the equality (or the bound) reveals that these expressions are not tight enough for real time data processing. This indicates that Theorem 4.4 provides only the initial ground work for an investigation of feedback control systems. However, this preparatory



1. $\underline{X} = D(\underline{Y} - \underline{V})$

2. $\underline{Z} = C(B(\underline{X}) + \underline{N})$

3. $\underline{V} = F(\underline{Z})$

4. $\underline{S} = B(D(\underline{Y}))$

Figure 4.1. The Disturbance Rejecting Feedback Control System (Regulator)

study is more than sufficient to demonstrate that it is again the Sensor Channel Transmittance that limits system performance. The developments in the future chapters are all based on this fundamental idea.

4.2 The Feedback Control Problem; Description

A recurring problem in the design of the class of automatic control systems, commonly referred to as regulators, is to isolate a dynamical system from undesirable disturbances. Typically this type of problem is solved through the use of feedback. A configuration, general enough to describe the most often encountered systems of this class is shown in Figure 4.1. Here y is the disturbing input and x is the system output. The sensor consists of a linear prefilter B, a postfilter C, and an equivalent noise source n, which is used to take into account any measurement error. The linear dynamical system of "plant" is given by D while the feedback gain is denoted by F. Often there are unchageable elements of the plant in the feedback loop. Since F is completely general it is convenient to assume that such elements are already included as part of the feedback signal processing. When the feedback gain, F, is chosen so as to achieve an optimum (in some sense) system output it will be written as $\hat{F}$.

When the sensor used in the feedback loop is absolutely accurate, i.e., there are no discretizing errors, measurement noise, distortion, etc., then the optimum system solution is to use a feedback network having infinite D.C. loop gain so that the resulting stable control system provides complete rejection of the disturbances. On the

other hand, if the sensor is very noisy when compared to the effects of the disturbing forces, then it would be expected that a low feedback gain would be the optimum solution.

Using entropy as a criterion function, it is obvious that values of feedback gain exist which will cause the system output to assume, continuously, all the values of entropy between the entropy of the sensor noise and the entropy of the uncontrolled disturbances. However, it is not so obvious that feedback can decrease the output entropy to less than either the signal entropy or the disturbance entropy. Therefore it is important to ask how much improvement (if any) over the uncontrolled system entropy can a given sensor provide when used in a feedback configuration. Since it is apparent that the performance of the resulting system is directly related to the properties of the sensor, it is desirable that the answer to this question should be based only on the invariant characteristics of the sensor. Theorem 4.4 below, shows exactly how the Channel Transmittance of the sensor is used to bound the performance improvement of a sampled data control system using the given sensor in a feedback loop.

4.3 System Uniqueness

The feedback system of Figure 4.1 may be described by the equation

$$\underline{W} = BD\underline{Y} + \underline{N} - BDF(C(\underline{W})) \ . \qquad (4.3.1)$$

This equation may be solved as a function of $BD\underline{Y}+\underline{N}$ to yield the expression

$$\underline{W} = g_1(BD\underline{Y}+\underline{N}) \; . \tag{4.3.2}$$

If g_1 is a single valued function of $BD\underline{Y}+\underline{N}$ it will be possible to write the informational inequality

$$I(\underline{n};\underline{W}) \leq I(\underline{n};BD\underline{Y}+\underline{N}) \tag{4.3.3}$$

for any arbitrary random vector $\underline{n}$. This is an important step in the development of entropy analysis of feedback systems and so it is critical to understand the conditions under which g_1 is a unique function of $BD\underline{Y}+\underline{N}$. To simplify the discussion somewhat, the signal is set equal to zero and only the noise response of the system is studied. The mathematical constraint on $g_1(\underline{N})$ is that the Jacobian

$$J(\underline{W},\underline{N}) = DET \left\{ \frac{\partial \underline{W}}{\partial \underline{N}} \right\} = DET\,[I+BDF_c'(\underline{W})\,]^{-1}$$

exist and be non-zero always with the same sign. For a physically realizable system this states that the perturbation response of the system is stable. This agrees with experience; the noise response of stable realizable systems is unique. The explanation of this lies in the fact that the effect of a noise input can't appear instantaneously at every point in the loop. One or more component outputs are not changed because of a noise input. Under the assumption that all system elements perform single valued transformations of the inputs it must follow that any signal in the loop is a single valued functional of the inputs and the initial conditions, if any.

When the feedback system is nonrealizable it is possible that the noise response is now unique. For example in Figure 4.1, take

$$D = d$$

$$B = 1$$

$$C = 1$$

$$F(z) = z^2$$

Then for a noise input, n,

$$w = n - dw^2$$

or w may have two values

$$w = \frac{-1 \pm \sqrt{1+4dn}}{2d}$$

but if w has two values so does x,

$$x = -d(x+n)^2$$

or

$$x^2 + x(2N + \frac{1}{d}) + n^2 = 0$$

$$x = - \frac{(2n + \frac{1}{d}) \pm \sqrt{(2n + \frac{1}{d})^2 - 4n^2}}{2}$$

This clearly indicates that the mathematical description of the system is incorrect, because what practical use can a multiple valued system output be when it is impossible to tell which output state will result from a given input? This type of system must be of limited use and will not be considered in this dissertation; only systems for which $g_1(\underline{N})$ is unique will be allowed.

4.4 The Entropy Theorem for Feedback Control

The entropy analysis of the regulator problem, Figure 4.1, is simplified by the use of the following lemma.

4.4.1 Lemma

For closed loop noise rejecting control systems of the form shown in Figure 4.1, with arbitrary sensor configuration, the joint entropy of the errors and the measurements is equal to the joint entropy of the open loop signal and the measurements, i.e.,

$$H(\underline{X},\underline{Z}) = H(\underline{U},\underline{Z})$$

where

$$\underline{U} = D(\underline{Y}).$$

For this lemma the plant $D(\underline{Y})$, is constrained to be linear.

Proof:

The variables $\underline{X}$ and $\underline{U}$ are related by

$$\underline{X} = D(\underline{Y}) - F(\underline{Z}).$$

Therefore

$$P_{x,z}(\underline{X},\underline{Z}) = P_{u,z}(\underline{X}+F(\underline{Z}),\underline{Z})$$

so that

$$H(\underline{X},\underline{Z}) = H(\underline{U},\underline{Z}).$$

Thus proving the lemma.

The results on an entropy analysis of the regulator of Figure 4.1 is summarized as Theorem 4.4. The main result is that the reduction

in the processed signal entropy, due to the use of feedback, cannot exceed the open loop Sensor Channel Transmittance.

4.4.2 Theorem 4.4

For the regulator problem shown in Figure 4.1, where B and D are linear, C is information preserving, and the feedback gain $F(\underline{Z})$ is arbitrary:

1. Minimizing the mutual information $I(\underline{X};\underline{W})$ is equivalent to minimizing the entropy of the error vector.

2. The entropy of the error vector, $H(\underline{X})$, always satisfies the inequality

$$H(\underline{X}) \geq H(\underline{U}) - I(\underline{U},\underline{S}+\underline{N}) = H_c \qquad (4.4.1)$$

$$\underline{U} = DY,$$

$$\underline{S} = BDY .$$

3. Since $I(\underline{U};\underline{S}+\underline{N})$ is the open loop Sensor Channel Transmittance it also follows that

$$H(\underline{U}) - H(\underline{X}) \leq H(\underline{S}+\underline{N}) - H(\underline{N}). \qquad (4.4.2)$$

The improvement in the system performance because of the use of feedback is bounded by an open loop quantity that is independent of $F(\underline{Z})$.

Proof:

The mutual information between the errors, $\underline{X}$, and the measurements, $\underline{Z}$, is

$$I(\underline{X},\underline{Z}) = H(\underline{X}) + H(\underline{Z}) - H(\underline{X},\underline{Z}) . \qquad (4.4.3)$$

Applying the lemma, and rewriting yields

$$I(\underline{X},\underline{Z}) = H(\underline{X}) - H(\underline{U}) + H(\underline{U}) + H(\underline{Z}) - H(\underline{U},\underline{Z})$$

$$= H(\underline{X}) - H(\underline{U}) + I(\underline{U};\underline{Z})$$

or

$$H(\underline{U}) - H(\underline{X}) = I(\underline{U};\underline{Z}) - I(\underline{X};\underline{Z}). \qquad (4.4.4)$$

This reuslt is not particularly useful since it utilizes $I(\underline{U};\underline{Z})$, which depends on the sensor closed loop Channel Transmittance. The reason that $I(\underline{U};\underline{Z})$ is not useful is that to calculate it, or measure it, requires specifying $F(\underline{Z})$, and the whole purpose of entropy analysis is to avoid doing that.

When the sensor noise is additive and the prefilter, B, is linear the closed loop Channel Transmittance may be related to the open loop Channel Transmittance as follows:

$$\underline{Z} = C(BD\underline{Y}+\underline{N} - BDF(\underline{Z})).$$

Using

$$\underline{S} \triangleq BD\underline{Y}$$

this equation may be solved to obtain $\underline{Z}$ as some function of $\underline{S}+\underline{N}$, i.e.,

$$\underline{Z} = g_2(\underline{S}+\underline{N}) = C(g_1(\underline{S}+\underline{N})).$$

As discussed in Section 4.3, g_2 must be single valued. Now, either g_2 preserves information, or it does not, but in either case,

$$I(\underline{U};\underline{Z}) \leq I(\underline{U};\underline{S}+\underline{N})$$

so that then

$$H(\underline{U}) - H(\underline{X}) \leq I(\underline{U};\underline{S}+\underline{N}) - I(\underline{X};\underline{Z}). \qquad (4.4.5)$$

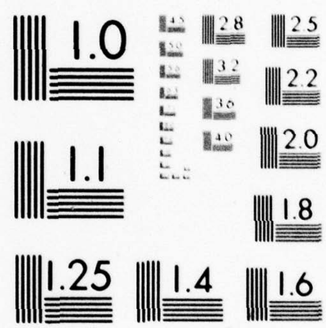
AD-A072 259 CALIFORNIA UNIV LOS ANGELES SCHOOL OF ENGINEERING A--ETC F/G 1/2
ENTROPHY ANALYSIS OF FEEDBACK FLIGHT DYNAMIC CONTROL SYSTEMS.(U)
JAN 79 H L WEIDEMANN, C T LEONDES F33615-77-C-3013

UNCLASSIFIED AFFDL-TR-78-123 NL

2 OF 2

AD A072259

END
DATE
FILMED
9-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Since only $H(\underline{X})$ and $I(\underline{X};\underline{Z})$ have a functional dependence on $F(\underline{Z})$ it must follow that minimizing $H(\underline{X})$ must be identical to minimizing $I(\underline{X};\underline{Z})$. Since $I(\underline{X};\underline{Z})$ is always non-negative the lower bound on being able to use a feedback function, $F(\underline{Z})$, to reduce $H(\underline{X})$, must be the Sensor Channel Transmittance, $I(\underline{U};\underline{S}+\underline{N})$.

Using equation (2.7.10)

$$I(\underline{U};\underline{S}+\underline{N}) = H(\underline{S}+\underline{N}) - H(\underline{N}),$$

equation (4.4.5) becomes

$$H(\underline{U}) - H(\underline{X}) \leq H(\underline{S}+\underline{N}) - H(\underline{N}) = r_o \qquad (4.3.5)$$

with equality if $g_2(\underline{S}+\underline{N})$ is an information preserving transformation and if $F(\underline{Z})$ is chosen to achieve

$$p(\underline{X},\underline{Z}) = p_x(\underline{X})\ p_z(\underline{Z}).$$

Thus, an inequality for the joint entropy of the error coordinates has been found which is true for any feedback element, F, and does not depend on knowing the feedback gain function in order to evaluate it. The constant r_o is a function of the parameters of the sensor and the statistical properties of the input signal and immediately sets the lower bound on the system entropy performance, irrespective of how sophisticated the feedback optimization is made.

Thus it has been proven that the effectiveness of a feedback loop in reducing system errors is determined by the ability of the sensor to transmit information. This is an entirely new way for looking at feedback control systems. It is the first time that the

entropy flow of signals around a feedback loop has been examined, and the first time that it has been proven that system performance is determined by the entropy handling capabilities of the components.

Rearranging the terms of equation 4.3.3 and using

$I(\underline{X};\underline{Z}) \geq 0,$

leads to

$H(\underline{X}) \geq H(\underline{Y}/\underline{W}).$

This resembles previously obtained conditional entropy expressions and it also sets a lower bound on the possible system error entropy. It is disappointing that this result can not bridge the gap to the real time Gaussian linear solution as it was able to do for the estimation problems. The reason is real time data processing considerations are such that $H(\underline{X})$ can not be caused to be equal to $H(\underline{Y}/\underline{W})$.

Actually the question of achieving the bound for physically realizable systems is the only disenchanting facet of this theorem. The requirement for equality in equation 4.4.5 is that $\underline{I}(\underline{X};\underline{W})$ (or $I(\underline{X};\underline{Z})$) be equal to zero. Of course, this is just the condition that the function $F(\underline{Z})$ be chosen so that the measurements $\underline{Z}$ are independent of the error $\underline{X}$. In a realizable closed loop control system the error is generated in real time because the only data that may be taken by the sensor is data about the errors and without data there cannot be errors, etc. Therefore it is impossible to cause past coordinates of the error vector to be indpendent of present or future measurements,

using a physically realizable filter, F, and $H(\underline{X})$ can not be equal to H_c. This disadvantage in no way impairs the validity of the bounds derived non-realizable systems but it does reduce the usefulness of the theorem. The truly useful result must have a tighter bound and it must be attainable. Obtaining this type of result is considered so important that the contents of Chapter Six will be devoted to deriving the real time solutions for both the feedback and the estimation problems and to examining the properties of that solution.

CHAPTER FIVE

THE TRACKING PROBLEM

5.1 Introduction

As already demonstrated in Chapter Four, the most significant application of entropy analysis is for the study of feedback control systems. Through the use of Channel Transmittance concepts, entropy analysis provides an entirely new interpretation of the benefits and limitations on the use of feedback to improve system performance. This interpretation is in complete agreement with conventional verbal descriptions of such systems but until now, it has not had a suitable mathematical foundation.

This chapter continues the study of entropy analysis by examining the performance of a closed loop tracking system. Results are again derived which demonstrate that the performance improvement of a tracking system utilizing a feedback loop over that of an open loop system is directly limited by the Channel Transmittance of the system sensor.

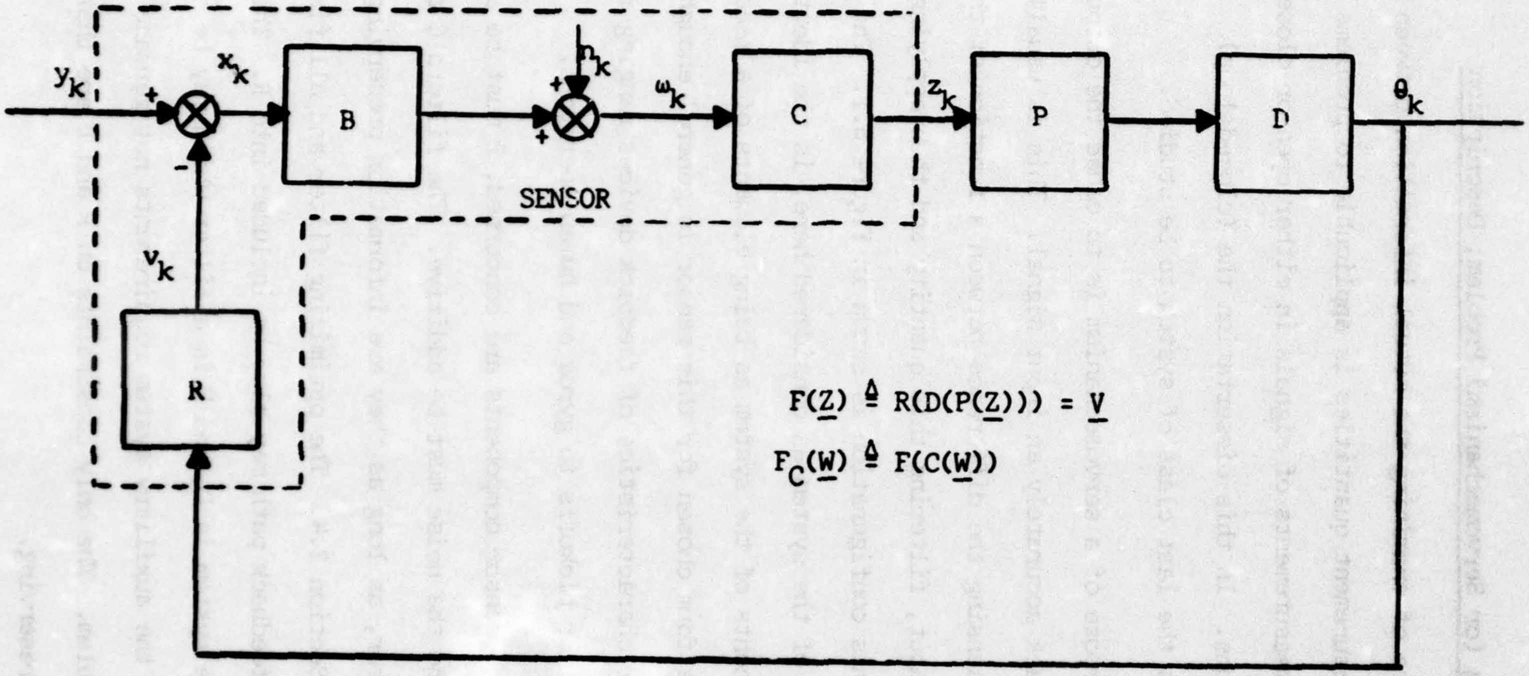
In this case the performance bounds can not be achieved and therefore are not as tight as would be desired. However the experience gained from the present limited approach paves the way for the sharper results presented in the next chapter and those results justify the continued emphasis of this dissertation on Channel Transmittance as a fundamental concept in estimation and feedback control.

5.2 The Tracking (or Servomechanism) Problem; Description

The procedure of examining the mutual information between the error and the measurement quantities is applicable to problems involving noisy measurements of signals in either open or closed loop configurations. In this dissertation the (closed loop) servomechanism is the last class of system to be studied.

The main purpose of a servomechanism is to cause the output to follow, or track accurately an input signal. This is usually accomplished by sensing the difference between a function of the output and the input, filtering this quantity and then applying it to the plant. This configuration is shown in Figure 5.1. The most important aspect of the system as considered here, is the identification of certain components of the system as being elements of a measurement device. The form chosen for this sensor is general enough to include the major characteristics of feedback devices ranging from simple potentiometer pickoffs to gyros and human operators.

As far as the sensor components are concerned, B must be a linear element and the noise must be additive. The filters C and R may be non-linear, as long as they are information preserving in the sense of Section 2.4. The optimizing filter and all fixed elements of the feedback path may also be included into R. The fixed part of the system is D, and P is a filter which may be added because of the auxiliary system requirements not germane to the tracking problem. The only constraints on P and D are that they be information preserving.



$$F(\underline{Z}) \triangleq R(D(P(\underline{Z}))) = \underline{V}$$

$$F_C(\underline{W}) \triangleq F(C(\underline{W}))$$

1. $\underline{X} = \underline{Y} - \underline{V}$
2. $\underline{Z} = C(B\underline{X}+\underline{N})$
3. $\underline{V} = F(\underline{Z})$

Figure 5.1. The Tracking Problem (Servomechanism).

5.3 The Entropy Theorem for Servomechanisms

The entropy analysis of the feedback tracking system is simplified through the use of the following lemma.

5.3.1 Lemma

For closed loop tracking systems of the form shown in Figure 5.1, having an arbitrary sensor configuration, the joint entropy of the signal and the measurements equals the joint entropy of the error and the measurements, i.e.,

$$H(\underline{X},\underline{Z}) = H(\underline{Y},\underline{Z}).$$

No further constraints are placed on the system components.

Proof:

$$\underline{X} = \underline{Y} - F(\underline{Z})$$

$$p_{x,z}(\underline{X},\underline{Z}) = p_{y,z}(\underline{X}+F(\underline{Z}),\underline{Z})$$

$$H(\underline{X},\underline{Z}) = H(\underline{Y},\underline{Z}) \qquad \text{Q.E.D.}$$

This lemma will now be used to prove the entropy theorem for servomechanisms. The major result of this theorem is that the improvement in system error performance because of the use of feedback is limited by the Sensor Channel Transmittance.

5.3.2 Theorem 5.3 (The Entropy Theorem for Servomechanisms)

For the tracking problem shown in Figure 5.1, where B is linear, C information preserving and the feedback gain, $F(\underline{Z})$ is arbitrary;

1. Minimizing the mutual information $I(\underline{X},\underline{W})$ is equivalent to minimizing the entropy of the error vector.

2. The entropy of the error vector, $H(\underline{X})$, always satisfies the inequality

$$H(\underline{X}) \leq H(\underline{Y}) - I(\underline{Y};B\underline{Y}+\underline{N}).$$

3. Since $I(y;B\underline{Y}+\underline{N})$ is the open loop Sensor Channel Transmittance it also follows that

$$H(\underline{Y}) - H(\underline{X}) \leq H(B\underline{Y}+\underline{N}) - H(\underline{N}).$$

Proof:

The proof for this theorem follows the same pattern as the proof for the regulator theorem (Section 4.4). The mutual information between the error $\underline{X}$ and the noisy measurement quantity $\underline{Z}$ is $I(\underline{X};\underline{Z})$, which may be written as

$$I(\underline{X};\underline{Z}) = H(\underline{X}) + H(\underline{Z}) - H(\underline{X},\underline{Z})$$

$$= H(\underline{X}) - H(\underline{Y}) + H(\underline{Y}) + H(\underline{Z}) - H(\underline{Y},\underline{Z})$$

$$H(\underline{Y}) - H(\underline{X}) = I(\underline{Y};\underline{Z}) - I(\underline{X};\underline{Z}).$$

$I(\underline{Y};\underline{Z})$ is the mutual information between the signal $\underline{Y}$ and the closed loop measurements $\underline{Z}$, and unless $F(\underline{Z})$ is given, it can not be calculated or measured. However, when the sensor is specifically constrained to have linear prefiltering and additive noise it follows that

$$\underline{Z} = C(B\underline{Y}+\underline{N} - BF(\underline{Z}))$$

$$= g_2(B\underline{Y}+\underline{N}) = g_1(C(B\underline{Y}+\underline{N}).$$

According to Section 4.3, $g_2(\cdot)$ is a single value function so

$$I(\underline{Y};\underline{Z}) \leq I(\underline{Y};B\underline{Y}+\underline{N}).$$

If $g_2(B\underline{Y}+\underline{N})$ is an information preserving transformation then this last equation is an equality.

$$H(\underline{Y}) - H(\underline{X}) \leq I(\underline{Y};B\underline{Y}+\underline{N}) - I(\underline{X};\underline{Z})$$

only $H(\underline{X})$ and $I(\underline{X};\underline{Z})$ are functions of $F(\underline{Z})$, so minimizing $I(\underline{X};\underline{Z})$ minimizes $H(\underline{X})$. Using

$$I(\underline{X};\underline{Z}) \geq 0$$

and

$$I(\underline{Y};B\underline{Y}+\underline{N}) = H(B\underline{Y}+\underline{N}) - H(\underline{N})$$

leads to the final equation

$$H(\underline{Y}) - H(\underline{X}) \leq H(B\underline{Y}+\underline{N}) - H(\underline{N}). \qquad \text{Q.E.D.}$$

CHAPTER SIX

REAL TIME DATA PROCESSING

6.1 Summary

An unfortunate limitation of the theorems of the preceding chapters is that, in order to achieve the lower bound, the processor must have access to the entire time history of the sensor output before it estimates the signal. This delay, which is required to make all possible information available, is not a serious defect for some categories of signal estimation, but it is disastrous in any real time estimation problem and catastrophic for feedback control systems where the presence of even small lags carries a high penalty.

Of course, this in no way impairs the validity of the bounds derived above, but their shortcomings due to the time lag requirement leads to the belief that in real time (or sequential) situations even tighter bounds may exist. If such a tightened bound exists, it is certainly obvious that the real time data processing system, where the information gained because of the new measurements can no longer be used to correct the previous errors, can have a performance no better than the total delay system, and probably has a poorer performance.

The work in this chapter develops the theory of sequential information and shows how this entirely new concept adapts itself readily to the real time pecularities of closed loop feedback systems. The first step is to define the Sequential Channel Transmittance of

the sensor. This property measures the sequential cumulative acquisition of information about the signal as measurements are made. This definition is contrived so that it completely discounts information derived from the new measurements about past signal samples. Of course this is what must be done in any real time processing system if it is to remain physically realizable.

The next step is to consider the Incremental Sequential Channel Transmittance, i.e., the amount of new information derived from one new measurement. The sum of all the incremental transmittances up to some time is obviously the Sequential Channel Transmittance. The incremental quantity meets two important needs. First, it measures the importance of new measurements. Secondly, if this quantity approaches zero, it provides an indication of what the asymptotic Sequential Channel Transmittance will be and how fast the asymptotic limit is approached. This asymptotic Sequential Transmittance is the irreducible uncertainty of the signal and no amount of additional measurements and data processing can decrease the conditioned signal entropy below this level, even if the data processing is optimum in the entropy sense.

Knowing that the total delay system performance of an estimation system is limited by Channel Transmittance as shown by Theorem 3.4, and suspecting that Sequential Channel Transmittance plays an analogous role in real time systems, it seems reasonable to assume that a theorem analogous to Theorem 3.4 can be proven for the real time estimation case. It turns out that such a theorem can be proven

by direct application of the following result, which will be proved as a lemma later.

"The joint entropy of the error and the measurements equals the joint entropy of signal and the measurements."

Similarly, for real time closed loop feedback systems, an analogous lemma and an analogous sequential theorem can be proven. For both the estimation and the feedback problem the entropy bounds are tight (i.e., achieveable) so that optimum realizable performance may be predicted and obtained. In the special Gaussian-linear case, the mathematical representations of the theorems for both types of systems admit simplifications which reduce them to the same results as obtained by conventional Gaussian-linear analysis, which shows the complete generality of the new channel approach to system analysis.

6.2 The Sequential Channel

In many situations it is desirable to know the Sequential Channel Transmittance of a sensor, which will be denoted by $I(y_k;\underline{Z})$ and which is defined by the relationship

$$I(y_k;\underline{Z}) = H(\underline{Z}) - H(\underline{Z}/y_k).$$

By continuing to interpret the sensor as a measurement device, it is seen that this quantity represents the information between the last input and all the previous measurements. It is easy to see that $I(y_k;\underline{Z}) \geq I(y_k;z_k)$, which is the information between the last input and the last output. This implies that even though some of the measurements z_i, $i \neq k$ were not made of the y_k variable, z_i is

is still statistically related to y_k and can therefore provide useful information about y_k.

The Sequential Channel Transmittance is different than the Channel Transmittance, $I(\underline{Y};\underline{Z})$, previously defined. In the previous case, Channel Transmittance is the information between all the inputs and all the measurements. This quantity implies that, in some way, time may be made to stand still until all the components of $\underline{Y}$ are measured, and then all the information in $\underline{Z}$ about $\underline{Y}$ may be utilized for whatever end is desired. Sequential Channel Transmittance implies just the opposite and emphasizes the inevitable march of time, and the uselessness of a measurement that comes too late.

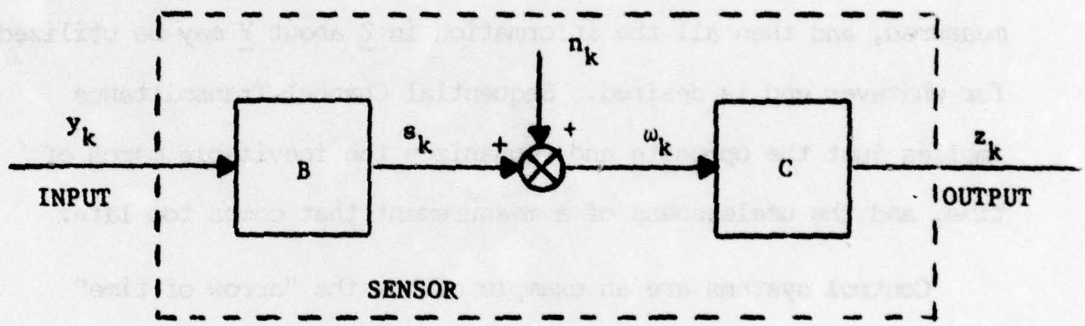
Control systems are an example of how the "arrow of time" renders some information useless, for no matter how much additional measurements are related (in a statistical sense) to previous inputs it is too late to make use of the information.

When C is an information preserving transformation, it follows that for the sensor shown in Figure 6.1

$$I(y_k;\underline{Z}) = I(y_k;\underline{W}) = H(\underline{W}) - H(\underline{W}/y_k).$$

since $H(\underline{W}) = H(B\underline{Y}+\underline{N})$, in order to determine $I(y_k;\underline{Z})$ it remains to determine the conditional entropy of $\underline{W}$ given y_k. A useful relationship between the respective probability density functions is

$$p_{wy}(\underline{W},\underline{Y}) = p_y(\underline{Y})\ p_n(\underline{W}-B\underline{Y}).$$



1. $\underline{S} = B(\underline{Y})$
2. $s_k = b_k(\underline{Y})$
3. $\underline{Z} = C(\underline{W})$
4. $z_k = c_k(\underline{W})$

Figure 6.1 A Typical Additive Noise Sensor.

Then

$$H(\underline{W}/y_k) \triangleq \int_{-\infty}^{\infty} d\underline{W} \int_{-\infty}^{\infty} dy_k \, p_{wy_k}(\underline{W},y_k) \, \mathrm{LOG} \, \frac{1}{p(\underline{W}/y_k)}$$

$$= \int_{-\infty}^{\infty} d\underline{W} \int_{-\infty}^{\infty} d\underline{Y} \, p_y(\underline{Y}) \, p_n(\underline{W}-B\underline{Y}) \, \mathrm{LOG} \, \frac{p(y_k)}{\int_{-\infty}^{\infty} d\underline{R}_{k-1} p_y(\underline{R}) p_n(\underline{W}-B\underline{R})}$$

where in the last integral the notation

$$r_k = y_k$$

and

$$d\underline{R}_{k-1} = (dr_{k-1})(dr_{k-2})\ldots(dr_1)$$

has been used.

6.3 Incremental Sequential Transmittance

The Incremental Sequential Transmittance is Δ_K, where the definition of this quantity is

$$\Delta_K \triangleq I(y_K;\underline{Z}_K) - I(y_{K-1};\underline{Z}_{K-1}).$$

In the stationary, time invarient case,

$$I(y_{K-1};\underline{Z}_{K-1}) \leq I(y_{K-1};\underline{Z}_{K-1}, z_o) = I(y_K;\underline{Z}_K)$$

therefore, $I(y_K;\underline{Z}_K)$ is a monotonically increasing function with K. Expand $I(y_K;\underline{Z}_K)$ according to property 16, Table I, Section 2.2,

$$I(y_K;\underline{Z}_K) = H(y_K) - H(y_K/\underline{Z}_K).$$

When the processes are stationary and y is generated by a markoff source,

$$H(y_K) \equiv H(y)$$

and, according to Theorem 2.2.2,

$$\lim_{K\to\infty} H(y_K/\underline{Z}_K) \quad H(y_\infty/\underline{Z}_\infty)$$

which implies that the Sequential Channel Transmittance approaches a steady-state value. Moreover, it follows that

$$\lim_{K\to\infty} \Delta_K = 0 \ .$$

The interpretation of this result is that even though the sequential information about y contained in the measurements is increasing, i.e., the entropy of y conditioned on the measurements is decreasing, the effectiveness of the oldest measurements to provide information about the latest signal is decreasing to zero.

6.4 The Sequential Entropy Theorem for Estimation

6.4.1 Introduction

The total vector derivations made in Chapter Three have hinted strongly that a real time entropy solution for the problem exists, and experience indicates that that solution will be a function of the real time capability of the sensor to transmit information, i.e., the Sensor Sequential Channel Transmittance property. But before this aspect of the problem is investigated, it is convenient to prove the following lemma which will be used in the proof of the sequential estimation theorem.

Lemma

For the signal estimating system shown in Figure 6.2, where C is information preserving, the joint entropy of the error and the measurements is equal to the joint entropy of the processed signal and the measurements, i.e.,

$$H(x_K,\underline{W}) = H(u_K,\underline{W}) \tag{6.4.1}$$

Proof:

The proof will proceed by examining $H(x_K,\underline{W})$ and $H(u_K,\underline{W})$ separately reducing the expressions so that they are a function of the fundamental quantities $p_y(\underline{Y})$ and $p_n(\underline{N})$, and then showing that the two expressions are indeed equal.

The formal definition of $H(x_K,\underline{W})$ is:

$$H(x_K,\underline{W}) = -\int_{-\infty}^{\infty} dx_K \int_{-\infty}^{\infty} d\underline{W}\; p(x_K,\underline{W}) \text{ LOG } p(x_K,\underline{W})$$

$$= -\int_{-\infty}^{\infty} d\underline{X} \int_{-\infty}^{\infty} d\underline{W}\; p_{xw}(\underline{X},\underline{W}) \text{ LOG } p(x_K,\underline{W}). \tag{6.4.2}$$

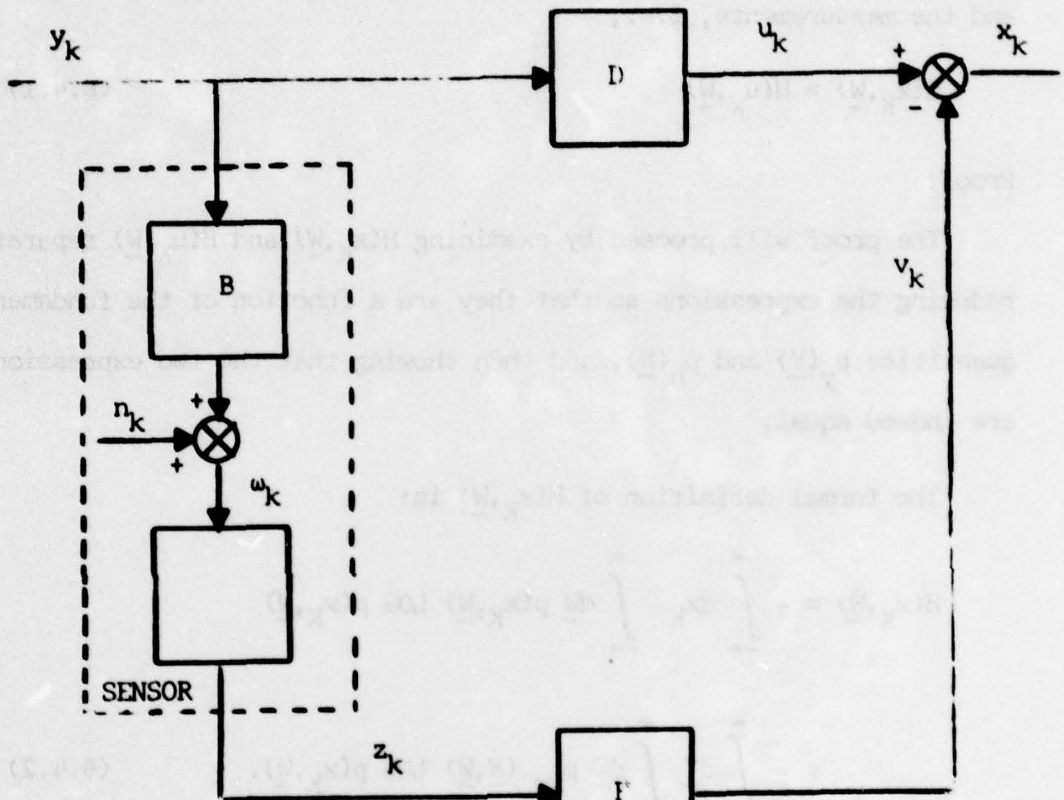
Using a previously derived expression for $p_{xw}(\underline{X},\underline{W})$, namely equation (3.4.9),

$$p_{xw}(\underline{X},\underline{W}) = p_y(\underline{Y})\, p_n(\underline{N})\, |\text{DET}\, [\partial_y\, D(\underline{Y})]|^{-1} \tag{6.4.3}$$

and the system equations

$$D(\underline{Y}) = \underline{X} + F_c(\underline{W}) \tag{6.4.4a}$$

$$\underline{N} = \underline{W} - BD^{-1}\, (\underline{X}+F_c(\underline{W}) \,. \tag{6.4.4b}$$



1. $\underline{U} = D(\underline{Y})$

2. $\underline{Z} = C(B(\underline{Y}) + \underline{N})$

3. $\underline{V} = F(\underline{Z})$

4. $\underline{X} = \underline{U} - \underline{V}$

Figure 6.2. The Sequential Estimation Problem.

It follows that:

$$[\partial_y D(\underline{Y})]^{-1} = \frac{\partial D^{-1}(\underline{X}+F_c(\underline{W}))}{\partial(\underline{X}+F_c(\underline{W})} \qquad (6.4.5)$$

and

$$p_{xw}(\underline{X},\underline{W}) = \left[p_y\left(D^{-1}\left(\underline{X}+F_c(\underline{W})\right)\right)\right]\left[p_n\left(\underline{W} - BD^{-1}\left(\underline{X}+F_c(\underline{W})\right)\right)\right]$$

$$\left|\mathrm{DET}\left[\frac{\partial D^{-1}\left(\underline{X}+F_c(\underline{W})\right)}{\partial\left(\underline{X}+F_c(\underline{W})\right)}\right]\right| . \qquad (6.4.6)$$

Using the probability formula and making the change of variable

$\underline{Y} = D^{-1}(\underline{X}+F_c(\underline{W})$

the joint entropy of the scalar x_K, and the vector $\underline{W}$ becomes

$$H(x_K,\underline{W}) = -\int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{W}\; p_y(\underline{Y})\; p_n(\underline{W}-B\underline{Y})\; \mathrm{LOG}\;[P]. \qquad (6.4.7)$$

where

$$[P] = \int_{-\infty}^{\infty} d\underline{X}_{K-1}\left[p_y\left(D^{-1}\left(\underline{X}+F_c(\underline{W})\right)\right)\right]\left[p_n\left(\underline{W}-BD^{-1}\left(\underline{X}+F_c(\underline{W})\right)\right)\right]$$

$$\left|\mathrm{DET}\left[\frac{\partial D^{-1}\left(\underline{X}+F_c(\underline{W})\right)}{\partial\left(\underline{X}+F_c(\underline{W})\right)}\right]\right| \qquad (6.4.8)$$

The important term in equation (6.4.7) (and also the most difficult term to handle) is the argument of the logarithm and it will now be studied separately.

The following simplifying vector notation is used:

$$\underline{X} = \text{COL}\,\{x_1, x_2, \ldots, x_K\} \tag{6.4.9a}$$

$$\underline{X}_{K-1} = \text{COL}\,\{x_1, x_2, \ldots, x_{K-1}\} \tag{6.49b}$$

$$x_K = d_K(\underline{Y}) - F_{c_K}(B\underline{Y}+\underline{N}) = d_K(\underline{Y}) - F_{c_K}(\underline{W}) \tag{6.4.9c}$$

$$d\underline{X}_{K-1} = dx_1\, dx_2 \ldots dx_{K-1} \tag{6.4.9d}$$

The quantity [P] can be simplified by making the change of variable:

$$\underline{R} = D^{-1}\,(\underline{X}+F_c(\underline{W})) \tag{6.4.10}$$

or using $\underline{X}$ as a function of $\underline{R}$ this transformation is:

$$\underline{X} = D(\underline{R}) - F_c(\underline{W})\,. \tag{6.4.11}$$

Then

$$[P] = \int_{-\infty}^{\infty} d\underline{R}_{K-1}\, p_y(\underline{R})\, p_n(\underline{W}-B\underline{R})\, \left| \text{DET}\,[\partial_r^K\, D(\underline{R})]^{-1}\, \text{DET}[\partial_r^{K-1}\, D(\underline{R}_{K-1})] \right|$$

(6.4.12)

where

$$\partial_y^K D = \begin{bmatrix} \dfrac{\partial d_1(\underline{Y})}{\partial y_1} & \cdots & \dfrac{\partial d_1(\underline{Y})}{\partial y_K} \\ \vdots & & \vdots \\ \dfrac{\partial d_K(\underline{Y})}{\partial y_1} & \cdots & \dfrac{\partial d_K(\underline{Y})}{\partial y_K} \end{bmatrix} \tag{6.4.13}$$

is the K dimensional Jacobian of the transformation $\underline{U} = D(\underline{Y})$ and

$$\partial_{y}^{K-1} D = \begin{bmatrix} \frac{\partial d_1(\underline{Y}_{K-1})}{\partial y_1} & \cdots & \frac{\partial d_1(\underline{Y}_{K-1})}{\partial y_{K-1}} \\ \vdots & & \vdots \\ \frac{\partial d_{K-1}(\underline{Y}_{K-1})}{\partial y_1} & \cdots & \frac{\partial d_{K-1}(\underline{Y}_{K-1})}{\partial y_{K-1}} \end{bmatrix} \qquad (6.4.14)$$

is the (K-1) dimensional Jacobian of the transformation $\underline{U}_{K-1}=D(\underline{Y}_{K-1})$ with $D(\underline{Y})$ defined as

$$D(\underline{Y}) = \begin{bmatrix} d_1(\underline{Y}) \\ \cdot \\ \cdot \\ \cdot \\ d_K(\underline{Y}) \end{bmatrix} .$$

When the simplified expression for [P] is used in the joint entropy $H(x_K,\underline{W})$ it becomes:

$$H(x_K,\underline{W}) = -\int_{-\infty}^{\infty} d\underline{Y}\, d\underline{W}\, p_y(\underline{Y})\, p_n(\underline{W}-B\underline{Y})\, \text{LOG}\left\{\int_{-\infty}^{\infty} d\underline{R}_{K-1}\, p_y(\underline{R})\, p_n(\underline{W}-B\underline{R}) \left|\text{DET}\,[\partial_r^K D]^{-1}\, \text{DET}\,[\partial_r^{K-1} D]\right|\right\} \qquad (6.4.15)$$

Having derived this expression, it is now convenient to shift attention and derive a similar (in fact an identical) expression for $H(u_K,\underline{W})$.

Consideration of $H(u_K,\underline{W})$ is based on examining the transformation between $[\underline{Y},\underline{N}]$ and $[\underline{W},\underline{U}]$, i.e.,

$$\underline{W} = B\underline{Y} + \underline{N}$$
$$\underline{U} = D(\underline{Y})$$

The Jacobian of this change of variables is given symbolically as:

$$J = |\mathrm{DET}\ \partial y\ D(\underline{Y})| = |\mathrm{DET} \left[\frac{\partial u}{\partial y} \right]| \tag{6.4.16}$$

so that the relationships between the various probability density functions are

$$p_{wu}(\underline{W},\underline{U})\ d\underline{W}\ d\underline{U} = p_y(\underline{Y})\ p_n(\underline{N})\ d\underline{Y}\ d\underline{N} \tag{6.4.17}$$

$$p_{wu}(\underline{W},\underline{U}) = p_y(D^{-1}(\underline{U}))\ p_n(\underline{W}-BD^{-1}(\underline{U}))\ |\mathrm{DET}\ \partial_u D^{-1}\ (\underline{U})|\ . \tag{6.4.18}$$

The joint entropy of u_K and $\underline{W}$, in terms of $p_y(\cdot)$ and $p_n(\cdot)$ is found by using the probability formulas in the entropy definition

$$H(u_K,\underline{W}) = -\int_{-\infty}^{\infty} d\underline{U} \int_{-\infty}^{\infty} d\underline{W}\ p_{uw}\ (\underline{U},\underline{W})\ \mathrm{LOG} \int_{-\infty}^{\infty} d\underline{Q}_{K-1}\ p_{uw}\ (\underline{Q},\underline{W}),$$

to yield

$$H(u_K,\underline{W}) = -\int_{-\infty}^{\infty} d\underline{U} \int_{-\infty}^{\infty} d\underline{W}\ |\mathrm{DET}\ [\partial_u^K\ D^{-1}(\underline{U})]|\ p_y(D^{-1}(\underline{U}))\ p_n(\underline{W}-BD^{-1}(\underline{U}))$$

$$\mathrm{LOG} \int_{-\infty}^{\infty} d\underline{Q}_{K-1}\ |\mathrm{DET}\ [\partial_q^K\ D^{-1}(\underline{Q})]|\ p_y(D^{-1}(\underline{Q}))p_n(\underline{W}-BD^{-1}(\underline{Q})) \tag{6.4.19}$$

where $\underline{Q}$ is a dummy variable of integration and $q_K=u_K$. It is important not to confuse the variables of the expectation integral with the variables of the integration performed within the LOG argument in order to achieve the necessary marginal probability distribution. In the expectation integral make the change of variable defined by:

$$D^{-1}(\underline{U}) = \underline{Y} \quad (6.4.20)$$

$$D^{-1}(\underline{Q}) = \underline{\alpha} \quad (6.4.21)$$

where $q_K = u_K = \alpha_K$.

With these new variables

$$d\underline{U} \; |\mathrm{DET}\; [\partial_u^K \; D^{-1}(\underline{U})]| = d\underline{Y} \quad (6.4.22)$$

$$[\partial_q^K \; D^{-1}(\underline{Q})] = [\partial_\alpha^K \; D(\underline{\alpha})]^{-1} \quad (6.4.23)$$

$$d\underline{Q}_{K-1} = d\underline{\alpha}_{K-1} \left|\mathrm{DET}\; [\partial_\alpha^{K-1} \; D(\underline{\alpha}_{K-1})]^{-1}\right| . \quad (6.4.24)$$

It then follows that:

$$H(u_K,\underline{W}) = \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{W} \; p_y(\underline{Y}) \; p_n(\underline{W}-\underline{BY})$$

$$\mathrm{LOG}\left[\int_{-\infty}^{\infty} d\underline{\alpha}_{K-1} \; p_y(\underline{\alpha}) \; p_n(\underline{W}-B\underline{\alpha}) \; \frac{\left|\mathrm{DET}\left[\partial_\alpha^{k-1} D(\underline{\alpha}_{K-1})\right]\right|}{\left|\mathrm{DET}\left[\partial_\alpha^{k} D(\underline{\alpha})\right]\right|}\right] \quad (6.4.25)$$

This expression is identical to the expansion for $H(x_K,\underline{W})$ therefore

$$H(u_K,\underline{W}) = H(x_K,\underline{W}) \quad (6.4.26)$$

and the lemma is proved.

6.4.2 Proof of the Theorem for Sequential Estimation

Theorem 6.4:

1. For the general estimation problem shown in Figure 6.2, with arbitrary $F(\underline{Z})$ and C information preserving, the entropy of single scalar error always satisfies the equality

$$I(x_K;\underline{W}) = H(x_K) + H(\underline{W}) - H(d_K\underline{Y},\underline{W}) \qquad (6.4.27)$$

and the inequality

$$H(x_K) \geq H(d_K\underline{Y}) + H(B\underline{Y}+\underline{N}/d_K\underline{Y}) - H(B\underline{Y}+N) = \tilde{H}_o \qquad (6.4.28)$$

where $\tilde{H}_o$ is indendent of $F(\underline{Z})$.

2. The system performance improvement due to feed-forward estimation is limited by the Sensor Sequential Channel Transmittance

$$H(d_K\underline{Y}) - H(x_K) \leq H(B\underline{Y}+\underline{N}) - H(B\underline{Y}+\underline{N}/d_K\underline{Y}) = I(d_K\underline{Y};B\underline{Y}+\underline{N}). \qquad (6.4.29)$$

3. Minimizing the mutual information $I(x_K;\underline{W})$ is equivalent to minimizing the error entropy.

4. The minimum error entropy occurs when $I(x_K;\underline{W}) = 0$ and is

$$H(x_K)\Big|_{min} = H_o = H(d_K\underline{Y}) + H(B\underline{Y}+\underline{N}/d_K\underline{Y}) - H(B\underline{Y}+\underline{N}) . \qquad (6.4.30)$$

This value is attained if the optimum filter $\hat{F}$ is chosen so that the most recent error, x_K, is independent of all the previous measurements, $\underline{W}$.

Proof:

The proof, although more involved than those studied previously, follows the format that has been successfully applied to the total

error vector case. As before, begin by examining the mutual information of the scalar x_K and the vector $\underline{W}$.

$$I(x_K;\underline{W}) = H(x_K) + H(\underline{W}) - H(\underline{W},x_K) \tag{6.4.31}$$

$H(\underline{W})$ is $H(B\underline{Y}+\underline{N})$, so the only problem is to determine $H(\underline{W},x_K)$,

$$H(x_K,\underline{W}) = -I(u_K;\underline{W}) + H(\underline{W}) + H(u_K) \equiv H(u_K,\underline{W}). \tag{6.4.32}$$

This last equation can now be introduced into the mutual information $I(x_K,\underline{W})$ to yield the basic result of this theorem,

$$I(x_K;\underline{W}) = H(x_K) + H(\underline{W}) + I(u_K;\underline{W}) - H(\underline{W}) - H(u_K)$$

$$I(x_K;\underline{W}) = H(x_K) - H(u_K) + I(u_K;\underline{W}) . \tag{6.4.33}$$

Rearranging this equation produces

$$H(u_K) - H(x_K) = I(u_K;\underline{W}) - I(x_K;\underline{W}) \tag{6.4.34}$$

and

$$H(u_K) - H(x_K) \leq I(u_K;\underline{W}) \tag{6.4.35}$$

and the theorem is proved.

The quantity $I(u_K;\underline{W})$ is the mutual information between a sample of the processed signal u_K and the measurements of $\underline{Y}$. It appears that the ability of feed-forward sequential estimation to reduce the uncontrolled entropy of a processed signal is limited by the amount of information provided about the processed signal due to measurements of the signal. This observation remains true even when the sensor does not have the specific form given in Figure 6.2. The proof of this is given in the following corollary.

Corollary I:

If a signal y_K, which is the last element of the sequence, $y_1, y_2, \ldots, y_k, \ldots, y_K$, is measured by a sensor having no other description other than its processed Channel Transmittance, $I(u_K;\underline{Z})$, then the entropy of the error $H(x_K)$, in estimating $d_K(\underline{Y})$ is given by

$$H(x_K) = H(d_K(\underline{Y})) + I(x_K;\underline{Z}) - I(d_K(\underline{Y});\underline{Z}) \;. \qquad (6.4.36)$$

If $f_K(\underline{Z})$ is chosen to achieve

$$I(x_K;\underline{Z}) = 0$$

then

$$\underset{f_K(\underline{Z})}{\text{MIN}} \{H(x_K)\} = \hat{H}(x_K) = H(d_K(\underline{Y})) - I(d_K(\underline{Y});\underline{Z}) \;. \qquad (6.4.37)$$

Note that neither $d_K(\underline{Y})$ or $f_K(\underline{Z})$ are constrained to preserve information.

Proof:

Figure 6.3 yields the system equations

$$x_K = d_K(\underline{Y}) - f_K(\underline{Z})$$

$$\underline{Z} = 0\underline{Y} + \underline{Z}$$

since

$$d_K(\underline{Y}) = u_K,$$

the joint probability density function of x_K and $\underline{Z}$ is

$$p_{xz}\,(x_K,\underline{Z}) = p_{uz}\,(x_K + f_K(\underline{Z}),\underline{Z}),$$

from which it follows that

$$H(x_K,\underline{Z}) = H(u_K,\underline{Z}) = H(d_K(\underline{Y}),\underline{Z}).$$

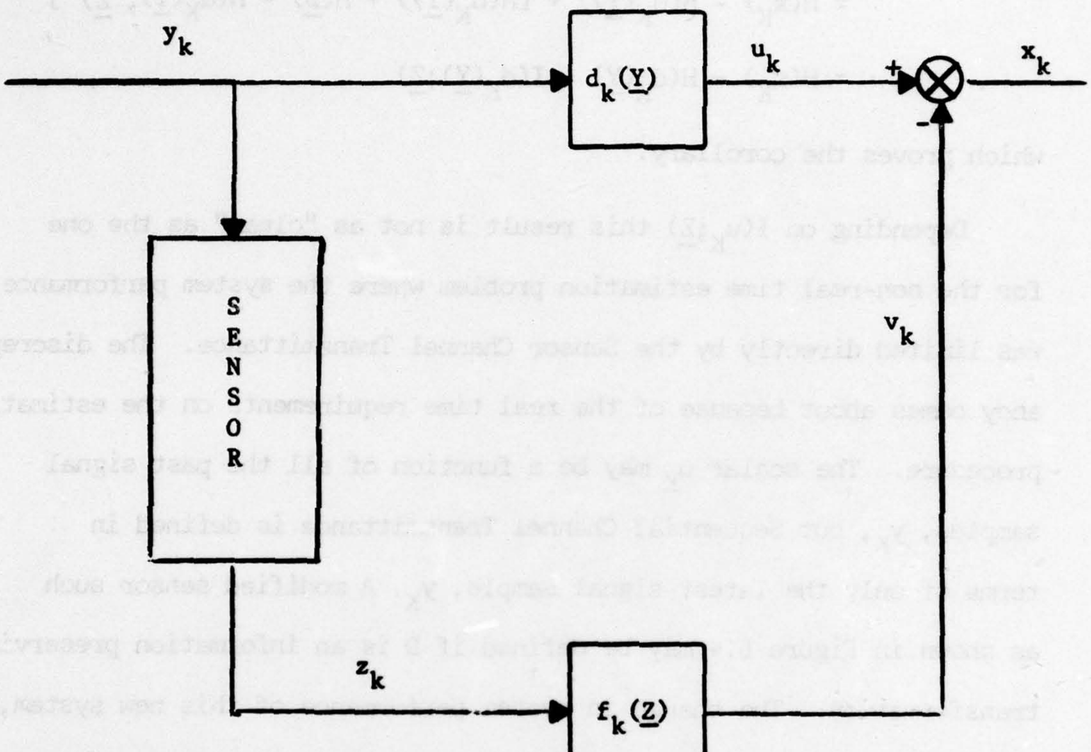


Figure 6.3. The Real Time Estimation Problem with Generalized Sensor.

This entropy equation is true regardless of the nature of $f_K(\underline{Z})$ and no constraints as to information preservation are implied. The mutual information between the error and the measurements is

$$
\begin{aligned}
I(x_K;\underline{Z}) &= H(x_K) + H(\underline{Z}) - H(x_K,\underline{Z}) \\
&= H(x_K) - H(d_K(\underline{Y})) + [H(d_K(\underline{Y})) + H(\underline{Z}) - H(d_K(\underline{Y}), \underline{Z})\ ] \\
&= H(x_K) - H(d_K(\underline{Y}) + I(d_K(\underline{Y});\underline{Z})
\end{aligned}
$$

which proves the corollary.

Depending on $I(u_K;\underline{Z})$ this result is not as "clean" as the one for the non-real time estimation problem where the system performance was limited directly by the Sensor Channel Transmittance. The discrepancy comes about because of the real time requirements on the estimating procedure. The scalar u_K may be a function of all the past signal samples, y_K, *but Sequential Channel Transmittance is defined in* terms of only the latest signal sample, y_K. A modified sensor such as shown in Figure 6.4 may be defined if D is an information preserving transformation. The change in system performance of this new system, due to feed-forward estimation, is now limited by the modified sensor Sequential Channel Transmittance.*

When no reliable analytical descriptions of the system components are available, the conclusion of Corollary I implies that it would be

*The modification of the system in this manner is not entirely analogous to the situation that results from taking D as the identity operator but bears a resemblance to that approach in that if D=I, channel capacity is again the performance limiting factor.

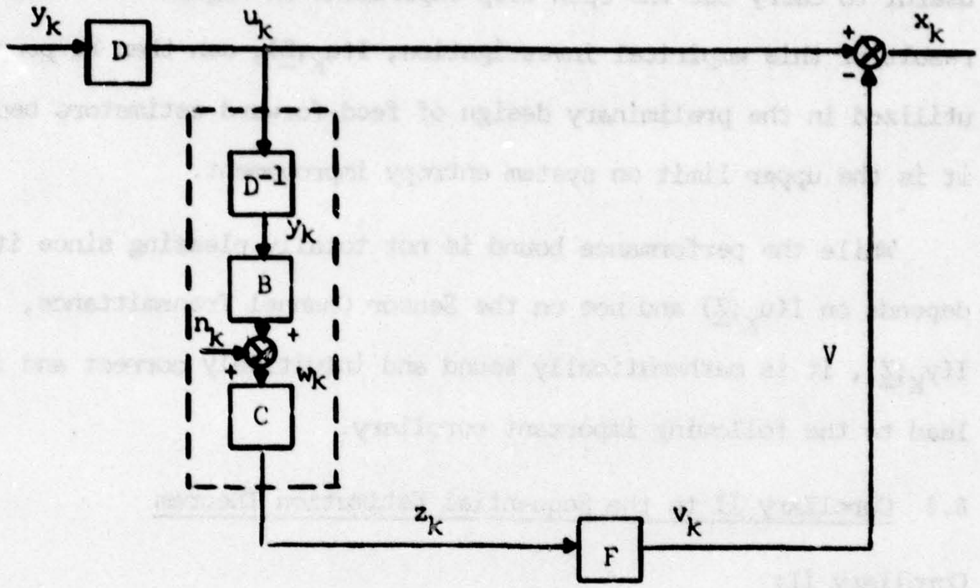


Figure 6.4. A Modified Estimation Problem.

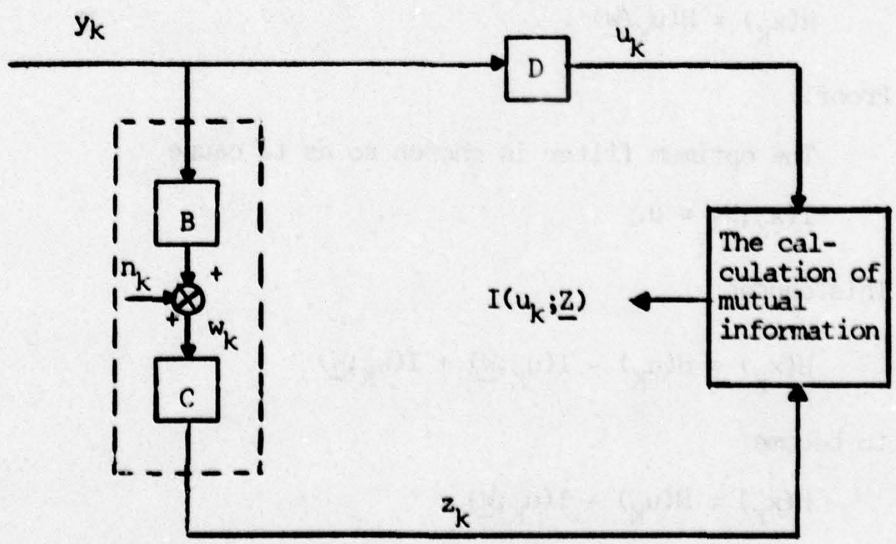


Figure 6.5. A Recommended Experiment.

useful to carry out the open loop experiment in Figure 6.5. The result of this empirical investigation, $I(u_K;\underline{Z})$, can then be properly utilized in the preliminary design of feed-forward estimators because it is the upper limit on system entropy improvement.

While the performance bound is not totally pleasing since it depends on $I(u_K;\underline{Z})$ and not on the Sensor Channel Transmittance, $I(y_K;\underline{Z})$, it is mathematically sound and intuitively correct and it does lead to the following important corollary.

6.5 Corollary II to the Sequential Estimation Theorem

Corollary II:

When the causal optimum filter is chosen so that the most recent error is independent of all measurements, the entropy of the error is equal to the conditional entropy of the processed signal (conditioned on all previous measurements), i.e.,

$$H(x_K) = H(u_K/\underline{W}) \ . \tag{6.5.1}$$

Proof:

The optimum filter is chosen so as to cause

$I(x_K;\underline{W}) = 0.$

This causes

$$H(x_K) = H(u_K) - I(u_K;\underline{W}) + I(x_K;\underline{W}) \tag{6.5.2}$$

to become

$$H(x_K) = H(u_K) - I(u_K;\underline{W}) \tag{6.5.3}$$

$$= H(u_K) - [H(u_K) - H(u_K/\underline{W})] \tag{6.5.4}$$

$$H(x_K) = H(u_K/\underline{W}) \qquad \text{Q.E.D.}$$

The uncertainty in x_K is the uncertainty of the processed input signal, given all the measurements.

Theorem 2.2.2 shows that when u_K is a markoff process and is stationary, $H(u_K/\underline{W})$ is monotonically decreasing and approaches a finite value and therefore

$$H(x_\infty) \leq H(x_K) \leq H(x_{K-1}) \qquad \forall\ K \tag{6.5.5}$$

Equation (6.5.5) is interesting because it demonstrates that under the stated conditions the error entropy is monotonically decreasing and approaches a finite steady-state value given as

$$H(x_\infty) = H(u_\infty/\underline{W}_\infty)$$

It is interesting to note that this result is not true in general. Under certain conditions the entropy of x can decrease without limit. For example, if the signal y were a constant then after a sufficiently long time the estimate of y derived from noisy measurements would have arbitrarily small error and the entropy of that error would have no lower bound.

6.6 Gaussian Example

The sample problem described in Section 3.6 is again examined. Contrary to the previous examples, the entropy solution to the real time Gaussian estimation of a signal in additive noise is not very simple. This is because the Gaussian mean square solution to the problem is not very simple.

First let us determine the minimum variance error using the classical techniques of Gaussian estimation theory. If y_K is the last

signal sample, then the best estimate of y_K is

$$v_K = \hat{F}\underline{Z}$$

where now $\underline{Z}$ is a column vector equal to $\underline{Y}+\underline{N}$ and $\hat{F}$ is the optimum row vector chosen so that the scalar quantity

$$\sigma^2 = E\,\{y_K - v_K\}^2$$

is minimized. It is well known that $\hat{F}$ will satisfy the orthogonality condition [42, p.218],

$$E\,\{(y_K - \hat{F}\underline{Z})(\underline{Z}^T)\} = 0$$

or

$$\hat{F}\,E\{\underline{Z}\underline{Z}^T\} = E\{y_K\underline{Z}^T\}\ .$$

Then using the following notation

$$R_y \triangleq E\{\underline{Y}\underline{Y}^T\}$$

$$\sigma_y^2 \triangleq E\{y_K^2\}$$

$$R_n \triangleq E\{\underline{N}\underline{N}^T\}$$

$$R_{yY} \triangleq E\{y_K\underline{Y}^T\}$$

for zero mean random variables. Now $\hat{F}$ is

$$\hat{F} = R_{yY}\,[R_y + R_n]^{-1}.$$

R_{yY} is a row vector and $[R_y+R_n]$ is a square matrix. The minimum variance error using this solution is σ^2.

$$\sigma^2 = E\,(y_K - \hat{F}\underline{Z})^2\} = E\{y_K^2 - y_K\hat{F}\underline{Z} - \hat{F}\underline{Z}y_K + \hat{F}\underline{Z}\underline{Z}^T\hat{F}^T\}$$

$$\sigma^2 = E\{y_K^2 + [\hat{F}\underline{Z}\underline{Z}^T - y_K\underline{Z}^T]\,\hat{F}^T - \hat{F}\underline{Z}y_K\}$$

$$\sigma^2 = E\{y_K^2\} - \hat{F}R_{yY}^T$$

$$\sigma^2 = \sigma_y^2 - R_{yY}\,[R_y + R_n]^{-1}\,R_{yY}^T\ .$$

It is obvious that

$$\sigma_y^2 = R_{yY} [R_y+R_n]^{-1} [R_y+R_n] \hat{\underline{I}}$$

where $\hat{\underline{I}}$ is the column vector with zero entries in all but the Kth coordinate, i.e.,

$$\hat{\underline{I}} = \mathrm{COL}\,\{0,0,\ldots,1\} .$$

Using this expression the minimum error variance becomes:

$$\sigma^2 = R_{yY} [R_y+R_n]^{-1} \{(R_y+R_n) \hat{\underline{I}} - R_{yY}^T\}$$

since

$$R_y\hat{\underline{I}} = R_{yY}^T$$

and

$$R_n \hat{\underline{I}} = R_{nN}^T \triangleq E\{n_K\underline{N}^T\}^T$$

$$\sigma^2 = R_{yY} [R_y+R_n]^{-1} R_{nN}^T .$$

Because x is a Gaussian random variable its entropy is:

$$H(x_K) = \frac{1}{2} \mathrm{LOG}\,(2\pi e\sigma^2) .$$

This is the final result and resembles very much, in form, the error found for the non-casual case in Section 3.6.

The entropy theorem on real time estimation applies, with the following system parameters:

$$D = I$$

$$B = I$$

$$C = I$$

$$\underline{Z} = \underline{Y}+\underline{N}$$

Since all the variables are Gaussian, the optimum filter is linear, the error can be made independent of the measurements and the equality

$$H(x_K) = H(y_K) + H(\underline{Y}+\underline{N}/y_K) - H(\underline{Y}+\underline{N})$$

is attained.

It is true that

$$H(\underline{Y}+\underline{N}/y_K) = H(\underline{Y}+\underline{N},y_K) - H(y_K)$$

so that

$$H(x_K) = H(\underline{Y}+\underline{N},y_K) - H(\underline{Y}+\underline{N}).$$

The first step in determining this solution is to evaluate $H(\underline{Y}+\underline{N},y_K)$.

$$H(\underline{Y}+\underline{N},y_K) = \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{N}\ p_y(\underline{Y})\ p_n(\underline{N})\ \mathrm{LOG}\ \frac{1}{\int_{-\infty}^{\infty} d\underline{Q}_{K-1}\ p_y(\underline{Q})p_n(\underline{Y}+\underline{N}-\underline{Q})}$$

where the equalities

$$\underline{Z} = \underline{Y}+\underline{N}$$

$$p_{y+n,y}(\underline{Q},\underline{Z}) = p_y(\underline{Q})\ p_n(\underline{Z}-\underline{Q})$$

and

$$y_K = q_K$$

have been used.

Then

$$p_y(\underline{Q}) = \frac{e^{-\frac{1}{2}\underline{Q}^T R_y^{-1}\underline{Q}}}{(2\pi)^{K/2}\sqrt{\mathrm{DET}\ [R_y]}}$$

$$p_n(\underline{Z}-\underline{Q}) = \frac{e^{-\frac{1}{2}(\underline{Z}-\underline{Q})^T R_n^{-1}(\underline{Z}-\underline{Q})}}{(2\pi)^{K/2}\sqrt{\mathrm{DET}\,[R_n]}}$$

$$p_y(\underline{Q})\, p_n(\underline{Z}-\underline{Q}) = \frac{1}{(2\pi)^K \sqrt{\mathrm{DET}\,[R_n R_y]}} \exp - \frac{1}{2}\{\underline{Q}^T R_y^{-1}\underline{Q} + (\underline{Z}-\underline{Q})^T R_n^{-1}\,(\underline{Z}-\underline{Q})\}$$

Now examine the exponent and expand

$$\underline{Q}^T R_y^{-1}\,\underline{Q} + (\underline{Z}-\underline{Q})^T R_n^{-1}\,(\underline{Z}-\underline{Q}) = \underline{Q}^T R_y^{-1}\underline{Q} + \underline{Z}^T R_n^{-1}\underline{Z} - \underline{Q}^T R_n^{-1}\underline{Z} - \underline{Z}^T R_n^{-1}\underline{Q} + \underline{Q}^T R_n^{-1}\underline{Q}$$

$$= \underline{Q}^T\,[R_y^{-1}+R_n^{-1}]\,\underline{Q} - \underline{Q}^T R_n^{-1}\underline{Z} - \underline{Z}^T R_n^{-1}\underline{Q} + \underline{Z}^T R_n^{-1}\underline{Z}$$

$$= [\underline{Q}^T - (R_y^{-1}+R_n^{-1})^{-1}\,R_n^{-1}\underline{Z}]^T\,[R_y^{-1}+R_n^{-1}]\,[\underline{Q} - (R_y^{-1}+R_n^{-1})\,R_n^{-1}\underline{Z}]$$

$$- \underline{Z}^T R_N^{-1}\,(R_y^{-1}+R_n^{-1})^{-1}\,[R_y^{-1}+R_n^{-1}]\,(R_y^{-1}+R_n^{-1})\,(R_n^{-1}\underline{Z}) + \underline{Z}^T R_n^{-1}\underline{Z}$$

$$= [\underline{Q}-R_y(R_y+R_n)^{-1}\underline{Z}]^T\,[R_y^{-1}+R_n^{-1}]\,(\underline{Q}-R_y(R_y+R_n)^{-1}\underline{Z}]$$

$$+ \underline{Z}^T\,[R_y+R_n]^{-1}\underline{Z}\,.$$

Using this last expression the term $p_y(\underline{Q})\ p_n(\underline{Z}-\underline{Q})$ may be written as the product of three terms, two which may be identified as being Gaussian probability density functions.

$$p_y(\underline{Q})\ p_n(\underline{Z}-\underline{Q}) = \frac{\sqrt{\mathrm{DET}\,[R_y+R_n]}}{\sqrt{\mathrm{DET}\,[R_y R_n]\ \mathrm{DET}\,[R_y^{-1}+R_n^{-1}]}}\ G_1 G_2$$

where G_1 and G_2 are Gaussian probability density functions defined as

$$G_1 = \frac{e^{-\frac{1}{2}[\underline{Q}-R_y(R_y+R_n)^{-1}\underline{Z}]^T [R_y^{-1}+R_n^{-1}] [\underline{Q}-R_y(R_y+R_n)^{-1}\underline{Z}]}}{\sqrt{(2\pi)^K \quad \text{DET}\,[R_y^{-1}+R_n^{-1}]^{-1}}}$$

and

$$G_2 = \frac{e^{-\frac{1}{2}\underline{Z}^T [R_y+R_n]^{-1}\underline{Z}}}{(2\pi)^{K/2}\sqrt{\text{DET}\,[R_y+R_n]}} = p_z(\underline{Z})$$

and DET $[R_yR_n]$ DET $[R_y^{-1}+R_n^{-1}]$ = DET $[R_y+R_n]$. Since G_2 is independent of $\underline{Q}$, and since the distribution of $\underline{Z} = \underline{Y} + \underline{N}$ is the more important one it follows that

$$H(\underline{Y}+\underline{N}, y_K) = \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{N}\; p_y(\underline{Y})\; p_n(\underline{N})\; LOG \frac{1}{\int_{-\infty}^{\infty} d\underline{Q}_{K-1}\, G_1}$$

$$+ \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{N}\; p_y(\underline{Y})\; p_n(\underline{N})\; \text{LOG}\, \frac{1}{p_z(\underline{Y}+\underline{N})}$$

Obviously the last term is $H(\underline{Y}+\underline{N})$ and the first term can be simplified as follows. G_1 is Gaussian and a function of the K variables

$q_1, q_2, \ldots, q_K$

$\int_{-\infty}^{\infty} d\underline{Q}_{K-1} G_1$ must also be Gaussian and is the function of only q_K, using

$\hat{\underline{I}}$ = COL $\{0,0,\ldots,1\}$ it may be expressed as:

$$\int_{-\infty}^{\infty} d\underline{Q}_{K-1} G_1 = \frac{e^{-\frac{1}{2}[q_K - \hat{\underline{I}}^T R_y (R_y+R_n)^{-1}\underline{Z}]^2 \, [\hat{\underline{I}}^T (R_y^{-1}+R_n^{-1}) \hat{\underline{I}}]}}{\sqrt{2\pi}\sqrt{\mathrm{DET}\,[\hat{\underline{I}}^T (R_y^{-1}+R_n^{-1})^{-1} \hat{\underline{I}}]^{-1}}}$$

Therefore the joint entropy $H(\underline{Y}+\underline{N},y_K)$ becomes

$$H(\underline{Y}+\underline{N},y_K) = H(\underline{Y}+\underline{N}) + \int_{-\infty}^{\infty} d\underline{Y} \int_{-\infty}^{\infty} d\underline{N}\; p_y(\underline{Y})\, p_n(\underline{N}) \left[\frac{\frac{1}{2}[y_K - \hat{\underline{I}}^T R_y (R_y+R_n)^{-1}\underline{Z}]^2}{\hat{\underline{I}}^T (R_y^{-1}+R_n^{-1})^{-1} \hat{\underline{I}}} \right]$$

$$+ \frac{1}{2} \mathrm{LOG} \left[2\pi\; \mathrm{DET}\,[\hat{\underline{I}}^T (R_y^{-1}+R_n^{-1})^{-1} \hat{\underline{I}}] \right].$$

In terms of the previous derivations

$$\mathrm{DET}\,[\hat{I}^T (R_y^{-1}+R_n^{-1})^{-1} \hat{\underline{I}}] = \mathrm{DET}\,[R_{yY}(R_y+R_n)^{-1} R_{nN}^T]$$

and

$$\hat{I}^T R_y (R_y+R_n)^{-1}\, \underline{Z} = R_{yY}(R_y+R_n)^{-1}\, \underline{Z}$$

moreover,

$$E\,\{[y_K - R_{yY}(R_y+R_n)^{-1}\, \underline{Z}]^2\} = \sigma_y^2 - 2R_{yY}(R_y+R_n)^{-1} R_{yY}^T + R_{yY}(R_y+R_n)^{-1} R_{yY}^T$$

$$= \sigma_y^2 - R_{yY}(R_y+R_n)^{-1} R_{yY}^T$$

$$= R_{yY}(R_y+R_n)^{-1} R_{nN}^T$$

so that finally,

$$H(\underline{Y}+\underline{N},y_K) = H(\underline{Y}+\underline{N}) + \frac{1}{2} + \frac{1}{2} \mathrm{LOG}\,[(2\pi)\, \mathrm{DET}\,[R_{yY}(R_y+R_n)^{-1} R_{nN}^T]\,].$$

since $\frac{1}{2} = \mathrm{LOG}_e$ (e), the error entropy at the Kth instant is now found to be

$$H(x_K) = \frac{1}{2}\,\mathrm{LOG}\left[(2\pi e)\ \mathrm{DET}\ [R_{yY}(R_y+R_n)^{-1}\ R_{nN}^T]\right]$$

which is, of course, identical to the solution obtained through the use of conventional mean square analysis.

6.7 The Sequential Entropy Theory for Feedback Control Systems

6.7.1 Introduction

The feedback problem has already been introduced in Chapter Four. In this section the concept of Sequential Channel Transmittance will be applied to the problem in order to determine a real time [physically realizable] solution. The following lemma will be a cornerstone of the Sequential Entropy Theorem for Feedback Control Systems.

Lemma

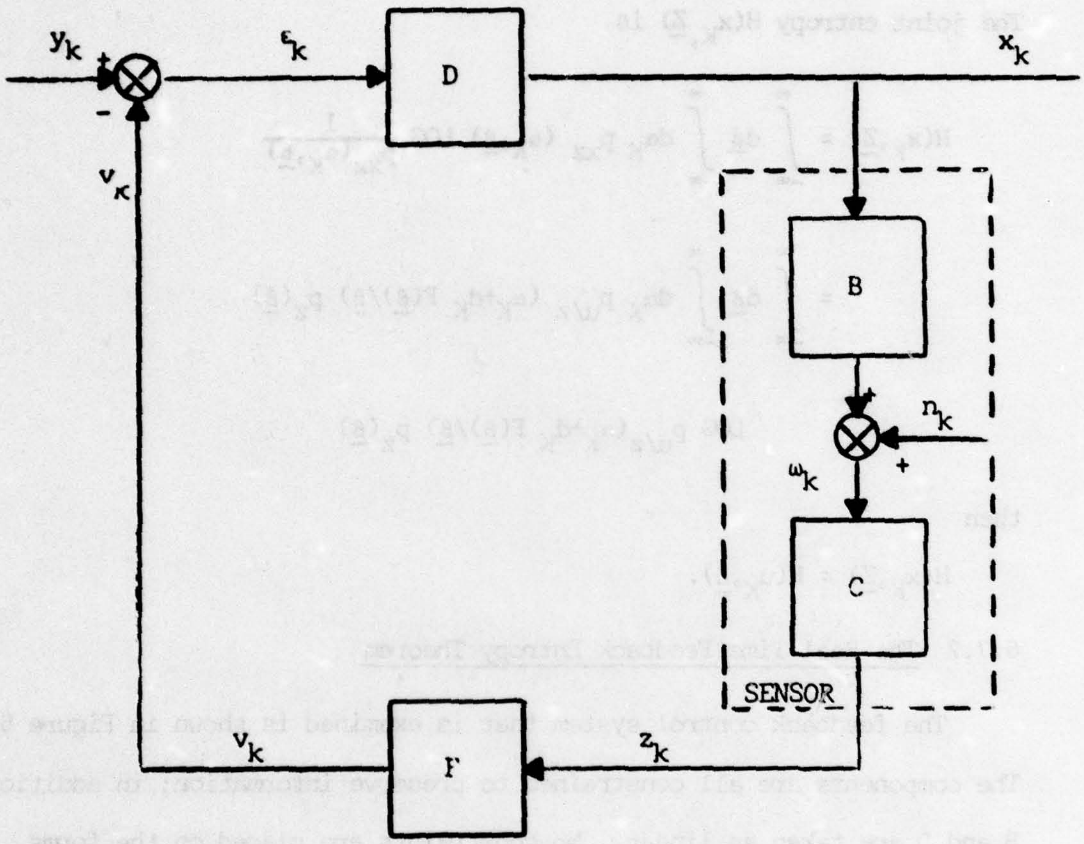
For the closed loop feedback control system of the form shown in Figure 6.6 with arbitrary system parameters and arbitrary sensor configuration, the joint entropy of the open loop signal and the closed loop measurements is identically equal to the joint entropy of the error and the closed loop measurements,

$$H(x_K,\underline{Z}) = H(u_K,\underline{Z}) \qquad (6.7.1)$$

Proof:

Let

$$D\underline{Y} = \underline{U} \qquad (6.7.2)$$



1. $\underline{X} = D(\underline{Y} - \underline{V})$
2. $\underline{Z} = C(B(\underline{X}) + \underline{N})$
3. $\underline{V} = F(\underline{Z})$
4. $\underline{S} = B(D(\underline{Y}))$

Figure 6.6. The Sequential Feedback Regulator.

then

$$x_K = u_K - d_K\, F(\underline{Z})$$

and

$$p_{x,z}\,(x_K,\underline{Z}) = p_{u,z}(x_K + d_K\, F(\underline{Z}),\underline{Z}).$$

The joint entropy $H(x_{K,}\underline{Z})$ is

$$H(x_K,\underline{Z}) = \int_{-\infty}^{\infty} d\underline{\beta} \int_{-\infty}^{\infty} d\alpha_K\, p_{xz}\,(\alpha_K,\underline{\beta})\ \mathrm{LOG}\ \frac{1}{p_{xz}(\alpha_K,\underline{\beta})}$$

$$= \int_{-\infty}^{\infty} d\underline{\beta} \int_{-\infty}^{\infty} d\alpha_K\, p_{u/z}\,(\alpha_K + d_K\, F(\underline{\beta})/\underline{\beta})\ p_z(\underline{\beta})$$

$$\mathrm{LOG}\ p_{u/z}(\alpha_K + d_K\, F(\underline{\beta})/\underline{\beta})\ p_z(\underline{\beta})$$

then

$$H(x_K,\underline{Z}) = H(u_K,\underline{Z}).$$

6.7.2 The Real Time Feedback Entropy Theorem

The feedback control system that is examined is shown in Figure 6.6. The components are all constrained to preserve information; in addition B and D are taken as linear. No constraints are placed on the forms of the signals other than the existence of an entropy measure. The following theorem is applicable.

Theorem 6.7

1. For any realization of the filter function $F(\underline{Z})$ the entropy of a single scalar error always satisfies the equality

$$I(x_K;\underline{Z}) = H(x_K) - H(u_K) + I(u_K;\underline{Z})\ , \qquad (6.7.3)$$

and the inequality

$$I(x_K;\underline{Z}) \leq H(x_K) - H(u_K) + I(u_K;\ B\underline{DY}+\underline{N}) \tag{6.7.3}$$

$\underline{U} = D\underline{Y}$

As a consequence, the following inequality is also true,

$$H(x_K) \geq (H(u_K) - I(u_K;B\underline{DY}+\underline{N}) = \tilde{H}_o \tag{6.7.4}$$

where $\tilde{H}_o$ is an open loop function and it is independent of $F(\underline{Z})$.

2. The presence of the term Sensor Sequential Channel Transmittance $(I(u_K;B\underline{DY}+\underline{N})$ in equation (6.7.4)) implies that the improvement in the system performance, because of the use of feedback, at least in the case of an additive noise sensor, must be limited by the open loop Sensor Sequential Channel Transmittance, i.e.,

$H(u_K) - H(x) \leq I(u_K;B\underline{DY}+\underline{N})$.

3. Minimizing the mutual information $I(x_K;\underline{Z})$ is equivalent to minimizing the entropy of the error.

4. If the filter function $\hat{F}(\underline{Z})$ is chosen so that the most recent error x_K, is independent of all the previous measurements $\underline{Z}$, then

$I(x_K;\underline{Z}) = 0$

and the minimum error entropy,

$$\underset{F(\underline{Z})}{\text{MIN}}\ \{H(x_K)\} = H_o = I(u_K;B\underline{DY}+\underline{N}) \tag{6.7.5}$$

is achieved.

Proof:

With the aid of the lemma of Section 6.7.1 the proof of these statements is quite simple. Begin with the mutual information

between the error x_K and the vector $\underline{Z}$

$$I(x_K;\underline{Z}) = H(x_K) + H(\underline{Z}) - H(x_K,\underline{Z})$$

$$= H(x_K) + H(\underline{Z}) - H(u_K,\underline{Z})$$

$$= H(x_K) - H(u_K) + I(u_K,\underline{Z}) .$$

Because this system is constrained to be physically realizable the equation

$$\underline{Z} = C(BD\underline{Y}+\underline{N}-BDF(\underline{Z}))$$

must have a unique single valued solution (see Section 4.3) given as

$$\underline{Z} = g_2(BD\underline{Y}+\underline{N}).$$

The function $g_2(\cdot)$ is a transformation of $BD\underline{Y}+\underline{N}$, the input, into $\underline{Z}$, the closed loop measurements. This transformation can not "create" information, therefore

$$I(u_K;\underline{Z}) \leq I(u_K;BD\underline{Y}+\underline{N}).$$

This leads to

$$I(x_K;\underline{Z}) \leq H(x_K) - H(u_K) + I(u_K;BD\underline{Y}+\underline{N}).$$

There is equality when the system transformation of $(BD\underline{Y}+\underline{N})$ into $\underline{Z}$ preserves information.

$I(u_K;BD\underline{Y}+\underline{N})$ is the open loop Sensor Sequential Channel Transmitance It is the information between the scalar signal u_K and the open loop measurement vector $BD\underline{Y}+\underline{N}$.

Since mutual information is non-zero, the inequality

$$H(x_K) \geq H(u_K) - I(u_K;BD\underline{Y}+\underline{N})$$

also follows. The achievement of the equality obviously corresponds to minimum error entropy and this must correspond to

$$I(x_K;\underline{Z}) = 0 . \qquad (6.7.6)$$

If u_K is interpreted as the system output with an open feedback path then

$$H(u_K) - H(x_K)$$

must be the system entropy improvement that results specifically from the use of feedback. It is truly interesting to note that this system improvement is bounded from above by the open loop sensor properties of the device that measures u_K. Without a doubt this is a radical new approach to feedback theory. This result is significant because it implies that optimum performance for feedback systems may be calculated without either

1. calculating the optimization network, or,
2. completing the feedback path and determining the closed loop signal entropies.

6.8 A Generalized Entropy Approach to Feedback Control

In the special case where B and D are linear and $g_2(\cdot)$ is information preserving it happens that

$$I(u_K;\underline{Z}) = I(u_K;BD\underline{Y})$$

but then it is also true that

$$H(u_K/\underline{Z}) = H(u_K/BD\underline{Y}).$$

If this equation were true for all types of sensors then the general, completely nonlinear, feedback problem would be solved.

It is not apparent at this time how the linear constraints on B and D in the feedback control problem may be relaxed. The following theorem is proposed as a step in that direction.

Theorem 6.8

If a sensor has the property that its closed loop conditional entropy equals the open loop conditional entropy, i.e., if

$$H(x_K/\underline{Z}) = H(u_K/\underline{Z}_o) \tag{6.8.1}$$

where $\underline{Z}$ is the closed loop output of the sensor and $\underline{Z}_o$ is the open loop output of the sensor when the input to the sensor is the vector $D\underline{Y}$, then

$$H(d_K\underline{Y}) - H(x_K) = I(\underline{Z}_o;u_K) - I(x_K;\underline{Z}) . \tag{6.8.2}$$

$I(\underline{Z}_o,u_K)$ is the open loop Sensor Sequential Channel Transmittance. In addition

$$H(u_K) - H(x_K) \leq I(\underline{Z}_o;u_K) . \tag{6.8.3}$$

Note: For the system studied in Section 6.7,

$$u_K = d_K \underline{Y}$$

and

$$\underline{Z}_o = BD\underline{Y}+\underline{N}.$$

Therefore it is obvious from the preceding work that at least the linear prefiltering sensor with additive noise satisfies this conditional entropy constraint, whether or not any significantly different sensors also satisfy this constraint is not known yet.

Proof:

$$I(x_K;\underline{Z}) = H(x_K) - H(x_K/\underline{Z})$$

$$= H(x_K) - H(u_K/\underline{Z}_o)$$

$$= H(x_K) - H(u_K) + H(u_K) - H(u_K/\underline{Z}_o)$$

$$= H(x_K) - H(u_K) - I(u_K;\underline{Z}_o)$$

or finally

$$H(u_K) - H(x_K) \leq I(u_K;\underline{Z}_o) \qquad \text{Q.E.D.}$$

Thus, under the assumptions of theorem 6.8 it is quite sufficient to consider the sensor as a device for transmitting information and its capability for doing so is described completely by the mutual information between the input and output. The actual form for the sensor model is immaterial. It seems reasonable to conjecture that this is true in general, but the proof of this conjecture has not been attained for arbitrary sensors.

In the real time feedback problem, the entropy solution may be shown to be very similar in form to results obtained for the Gaussian-linear system. This is stated as a corollary to theorem 6.7.

6.8.1 Corollary to Theorem 6.7

For any feedback system, having arbitrary components and unconstrained sensor models, the error entropy is always bounded by the conditional entropy of the error given the measurements, i.e.,

$$H(x_K) \geq H(u_K/\underline{Z}) \qquad (6.8.4)$$

The equality holds if and only if

$$I(x_K;\underline{Z}) = 0 \text{ .}$$

Proof:

Equation (6.7.3) was derived without any constraints being imposed on the system elements,

$$I(x_K;\underline{Z}) = H(x_K) - H(u_K) + I(u_K,\underline{Z}). \qquad (6.8.5)$$

Rearranging this equation yields

$$H(x) = H(u_K) - H(u_K) + H(u_K/\underline{Z}) + I(x_K;\underline{Z})$$

since

$$I(x_K;\underline{Z}) \geq 0$$

the corollary is proven.

CHAPTER SEVEN

CONTINUOUS TIME DATA PROCESSING

7.1 Discussion of the Difficulties in Solving Time Continuous Systems with Entropy

Barring consideration of the fact that so far useful results for the calculation of time continuous mutual information have only been obtained when the processes involved are Gaussian, the extension of entropy analysis, as described in the previous chapters of this dissertation, to time continuous systems is simultaneously easy and difficult. It is easy because all of the theorems in the continuous case are proven using only the properties of mutual information and these properties should hold in the continuous as well as the discrete case. It is difficult because there does not exist a fundamental understanding of continuous time entropy. Undoubtedly, the poor understanding of the properties of this type of entropy is due in no small part to the fact that up until now there has been no practical need for a continuous time entropy measure. Probably the spectre of infinite entropy has so intimidated researchers that they have not even made strong attempts to find useful applications. The following very simple example opens up the possibility of just such a useful application. The reader is cautioned to realize that, since only Gaussian variables can be studied, the results only indicate a potential on the part of entropy to solve certain broader time continuous problems.

The problem considered is the estimation of the magnitude of a Gaussian random D. C. signal in additive stationary Gaussian white noise. For this elementary problem the entropies of all the individual signal quantities exist, so that it would be no great difficulty to find the entropy of the estimation error, term by term, from the familiar equation:

$$H(x(t)) = H(Dy(t)) + H(n(t)) - H(By(t)+n(t)). \qquad (7.1.1)$$

However, realistic situations exist where some or all of these entropies can not be found. An alternate approach is suggested by the technique used in the proof of Theorem 3.5. There the concept of Channel Transmittance is utilized to obtain equation (3.5.3), rewritten here for time continuous processes,

$$H(Dy(t)) - H(x(t)) \leq I(y(t);z(t)) \qquad (7.1.2)$$

and since $I(y;z)$ almost always exists and can be found, it therefore follows that the bound in improvement in the error entropy uncertainty for any continuous estimating system almost always may be calculated. Of course, no absolute measure of the system entropy performance can now be obtained since $H(x)$ can not be considered as a reliable absolute measure in continuous time situations, and no other absolute measure of the error will result from analysis.

7.2 Example

Consider the D. C. signal

$$s(t) = a, \qquad 0 \leq t \leq T, \qquad (7.2.1)$$

"a" is a zero mean random variable that is constant over the specified time interval and it has a variance

$$\text{var}\,\{a\} \triangleq \sigma^2. \qquad (7.2.2)$$

Estimates of "a" are to be made using the noisy measurements z(t),

$$z(t) = a + n(t), \quad 0 \leq t \leq T,$$

the random signal n(t) is a zero mean stationary white noise process independent of "a" and having the covariance function

$$E\,\{n(t_1)\; n(t_2)\} = N_o\; \delta(t_2 - t_1). \qquad (7.2.3)$$

If "a" is estimated using

$$\hat{a} = \int_0^T h(T-\tau)\; z(\tau)\; d\tau\,. \qquad (7.2.4)$$

Then according to the Wiener-Hopf equation [42, p.408], the optimum estimating filter must satisfy

$$E\,\{a^2\} = \int_0^T h(T-\tau)\; E\{z(\tau)\; z(t)\}\; d\tau, \qquad (7.2.5)$$

and therefore

$$\sigma^2 = \sigma^2 \int_0^T h(T-\tau)\; d\tau + N_o\; h(T-t) \quad \forall\, t \epsilon\, [0,T]. \qquad (7.2.6)$$

The only filter function h(T-τ) that solves this equation is a constant, i.e.,

$$h(T-\tau) = \frac{\sigma^2}{N_o + \sigma^2 T}\,. \qquad (7.2.7)$$

After combining equation (7.2.7) with equation (7.2.4) the optimum estimate for "a" is the weighted integral of the measurements, i.e.,

$$\hat{a} = \int_0^T \frac{\sigma^2}{\sigma^2 T + N_o} \; z(\tau)\, d\tau \; . \tag{7.2.8}$$

The variance of the error in estimating "a" is found as follows:

$$E\,\{(a-a)^2\} = E\left\{a - \frac{\sigma^2}{\sigma^2 T + N_o}\int_0^T z(\tau)\, d\tau\right\}^2$$

$$= \sigma^2 - 2\,\frac{\sigma^2}{\sigma^2 T + N_o}\,(\sigma^2 T) + \left(\frac{\sigma^2}{\sigma^2 T + N_o}\right)^2 \left(\sigma^2 T^2 + N_o T\right)$$

$$= \sigma^2 - \frac{2(\sigma^2)^2 T}{\sigma^2 T + N_o} + \frac{(\sigma^2)^2 T}{\sigma^2 T + N_o}$$

$$= \frac{\sigma^2 N_o}{\sigma^2 T + N} \; . \tag{7.2.9}$$

Using the properties of Gaussian random variables the entropy of the estimation error must be

$$H(a-\hat{a}) = \frac{1}{2}\,\text{LOG}\left(\frac{\sigma^2 N_o}{\sigma^2 T + N_o}\; 2\pi e\right) . \tag{7.2.10}$$

This, taken together with the entropy of "a"

$$H(a) = \frac{1}{2}\,\text{LOG}\,(\sigma^2\; 2\pi e), \tag{7.2.11}$$

proves that the change in the entropy of "a," because an estimation is made is

$$H(a) - H(a-\hat{a}) = \frac{1}{2} \text{LOG} \left(\frac{N_o + \sigma^2 T}{N_o} \right) . \tag{7.2.12}$$

The same result may be obtained from a direct application of the Estimation Theorem, 3.5 (applied to continuous time systems and using the work of Hyang [33,p.59]).

According to Hyang,

$$I_T(s(t)+n(t);s(t)) = I_T(a+n(t);a) = \frac{1}{2} \sum_K \text{LOG}\ (1+\lambda_k), \tag{7.2.13}$$

where the subscript T denotes that the time interval over which the processes are defined is [0,T]. The λ_K found from the projection of the signal onto an eigenfunction space spanned by the set of functions $F_K(\tau)$, i.e.,

$$E\ \{\langle F_K,\ s\rangle^2\} = \lambda_K , \tag{7.2.14}$$

where the inner product is defined as

$$\langle F_K,\ s\rangle \triangleq \int_0^T s(\tau)\ F_K(\tau)\ d\tau . \tag{7.2.15}$$

The eigenfunctions F_K satisfy the integral equation

$$\lambda_K \int_0^T E\ \{n(t_1)\ n(t_2)\}\ F_K(t_2)\ dt_2 = \int_0^T E\ \{a^2\}\ F_K(t_2)\ dt_2. \tag{7.2.16}$$

For the simple D.C. process in white noise the only eigenfunction, F, is a constant, and

$$\lambda_1 = \frac{\sigma^2 T}{N_o}$$
$$\lambda_K = 0 \qquad K = 2, 3, \ldots \tag{7.2.17}$$

Then by equations (7.2.13) and (7.2.17)

$$I_T(a+n(t);a) = \frac{1}{2} \text{LOG} \left(1 + \frac{\sigma^2 T}{N_o}\right), \tag{7.2.18}$$

which, of course, completely agrees with the informational quantity of equation (7.2.12), which was calculated directly from mean square error analysis.

7.3 Conclusions

There is no doubt that the above information theoretic procedure may be applied to more complex time-continuous processes, but only if all the signals are Gaussian and their covariance functions have Karhunen-Loeve expansions. Fortunately one can expect that the Gaussian restriction will be removed with further research, so the only real problem to consider is whether equation (7.1.2), (repeated here)

$$H(Dy(t)) - H(x(t)) = I(y(t);z(t)) \tag{7.2.19}$$

has any meaning, and not whether the information, $I(y(t);z(t))$, can be calculated. In this context, $I(y(t);z(t))$ still retains its interpretation as a Channel Transmittance so that even in the continuous case it is the Sensor Channel Transmittance that governs the ability of the system to improve the entropy uncertainty of the signal. Equation (7.2.19) is effective for comparing the performances of different sensors used in optimum configurations. Unlike the equations for discrete time, the continuous entropy bounding equations can not be rewritten as variance bounding equations. This is because there

is no relationship between individual signal entropy [which is often undefined] and variance [which must be defined in the context of a continuous time random variable].

CHAPTER EIGHT

ENTROPY ANALYSIS OF ADAPTIVE CONTROL

8.1 Introduction

Underlying the whole theory of adaptive control is the idea that the output of the vaguely described system can be examined in order to obtain a better description of the system and ultimately to use this information to improve the control of the system. Implied in this procedure is the concept that the system output contains information about the system parameters which might be profitably used to estimate those parameters. If the experience gained in this dissertation for signal estimation is any indication, it should follow that entropy is a logical tool for describing parameter estimation. In the example presented below, a very simple situation is examined. The system output is a linear function of the system parameters which are to be estimated. This is not the typical type of problem encountered in adaptive control, however this model does represent a system with unknown initial conditions and could still be valuable. A solution is obtained by imbedding this problem into the estimation problem and applying the estimation Theorem 3.4. Even though severe restrictions are encountered in the useful application of entropy to the adaptive control problem, it is not overly optimistic to expect that they will soon be relaxed. In any case, the important conclusion is that the amount of improvement in the entropy uncertainty of a system parameter is limited by

the properties of the sensor measuring the output and the properties of the output function that is measured.

8.2 A Parameter Estimation Problem

Consider the following problem: The scalar output of a discrete time system at time t_i (i^{th} instant) is

$$w_i = \underline{M}_i^T \; \underline{\alpha} + n_i \qquad i = 1, \ldots, K$$

where $\underline{M}_i$ is a column vector, n_i is additive noise independent of $\underline{\alpha}$, and $\underline{\alpha}$ is a column vector of size K of parameters upon which the output depends linearly. It is desired to estimate $\underline{\alpha}$. This problem immediately falls into the context of the estimation problem, Section 3.4, especially when the following corresponds at the Kth instant are noted:

$$\begin{aligned}
\underline{Y} &= \underline{\alpha} \\
B &= \{\underline{M}_1, \underline{M}_2, \ldots M_K\}^T \\
C &= I \\
D &= I \\
\underline{Z} &= B\,\underline{\alpha} + \underline{N} \\
\underline{U} &= \underline{\alpha} \\
\underline{V} &= \underline{\hat{\alpha}} \\
\underline{X} &= \underline{\alpha} - \underline{\hat{\alpha}}
\end{aligned}$$

It then follows that the improvement in the entropy uncertainty of the system parameters is limited directly by the Channel Transmittance property of the sensor. This is a new approach to adaptive control analysis.

According to the results of the estimation theorem, the entropy of the error in the estimation of $\underline{\alpha}$ is bounded by

$$H(\underline{X}) \geq H(\underline{\alpha}) + H(\underline{N}) - H(B\underline{\alpha}+\underline{N})$$

where $\underline{\alpha}$ and $\underline{N}$ are zero mean Gaussian random vectors with covariance matrices

$$E\{\underline{\alpha}\,\underline{\alpha}^T\} = R_\alpha$$

$$E\{\underline{N}\,\underline{N}^T\} = R_n.$$

The entropy equation becomes

$$\begin{aligned} H(\underline{X}) \geq\ & \frac{1}{2} \text{LOG } \{(2\pi)^K \text{ DET } [R_\alpha]\} \\ & + \frac{1}{2} \text{LOG } \{(2\pi)^K \text{ DET } [R_n]\} \\ & - \frac{1}{2} \text{LOG } \{(2\pi)^K \text{ DET } [BR_\alpha B^T+R_n]\}, \end{aligned}$$

or, when using optimum estimates

$$H(\underline{X}) = \frac{1}{2} \text{LOG } \{(2\pi)^K \quad \frac{\text{DET } [R_\alpha] \text{ DET } [R_n]}{\text{DET } [BR_\alpha B^T+R_n]}$$

This same result may be obtained in the following manner from a point of view based only on mean square error analysis. If the estimation of $\underline{\alpha}$ takes the form

$$\hat{\underline{\alpha}} = A_o\, \underline{Z}$$

according to the principal of orthogonality, [42, p.218],

$$E\{(\underline{\alpha}-\hat{\underline{\alpha}})\, \underline{Z}^T\} = 0$$

or A_o must satisfy

$$R_\alpha B^T = A_o(BR_\alpha B^T+R_n).$$

Using the value of A_o as

$$A_o = [R_\alpha B^T]\,[BR_\alpha B^T+R_n]^{-1}$$

it is found directly that

$$\begin{aligned}
\mathrm{VAR}\,\{\underline{\alpha}-\hat{\underline{\alpha}}\} &= E\,\{(\underline{\alpha}-A_o\underline{Z})(\underline{\alpha}-A_o\underline{Z})^T\} \\
&= R_\alpha - A_o BR_\alpha \\
&= R_\alpha - (R_\alpha B^T)(BR_\alpha B^T+R_n)^{-1}\,BR_\alpha \\
&= R_\alpha\,[1-B^T(BR_\alpha B^T+R_n)^{-1}\,BR_\alpha] \\
&= R_\alpha B^T\,(BR_\alpha B^T+R_n)^{-1}\,[(BR_\alpha B^T+R_n)\,B^{-T}-BR_\alpha] \\
&= R_\alpha B^T\,(BR_\alpha B^T+R_n)^{-1}\,R_n B^{-T}
\end{aligned}$$

It follows from the constraint that $\mathrm{DET}[B] = \mathrm{DET}[B^T] = \mathrm{DET}[B^{-T}] \neq 0$, that

$$H(\underline{\alpha}-\hat{\underline{\alpha}}) = \frac{1}{2}\,\mathrm{LOG}\left\{(2\pi)^K\,\frac{\mathrm{DET}[R_\alpha]\,\mathrm{DET}[R_n]}{\mathrm{DET}[BR_\alpha B^T+R_n]}\right\}$$

This example demonstrates a possible potential on the part of entropy to be a useful analysis tool for adaptive control problems. However in this very simple example there is a significant limitation that requires further research to overcome. The significant limitation is the form of the constrant on the operation B, i.e.,

$$\mathrm{DET}\left[\frac{dB(\underline{Y})}{d\underline{Y}}\right] \neq 0$$

This restriction on the Jacobian of the transformation $B(\underline{Y})$ implies two conditions:

1. B is K valued function of K variables.

2. B preserves information.

Condition 1 insists that exactly K measruements be made if K parameters are to be estimated. This is not a common condition since most often redundant measurements are taken. Constraining $B(\underline{Y})$ in this manner was necessary for the proof of Theorem 3.4. However, this is not a requirement for the proof of Theorem 8.3 below, which allows both redundant measurements and nonlinear plant outputs.

8.3 Entropy Solution of the Identification Problem

Theorem 8.3

The entropy of the error, $\underline{X}_m$, in identifying an m dimensional parameter set, $\underline{\beta}_m$, satisfies

$$H(\underline{X}_m) \geq H(\underline{\beta}_m) - I(\underline{\beta}_m;\underline{Z}_K) + I(\underline{X}_m;\underline{Z}_K)$$

where the identification system has the form shown in Figure 8.1.

In general, the output signal of the plant to be identified is a nonlinear function of the parameters $\underline{\beta}_m$. This nonlinear operation has been included in the sensor model, so that $I(\underline{\beta}_m;\underline{Z}_K)$ is recognized to be the Channel Transmittance of the sensor. Obviously the improvement of the parameter entropy from $H(\underline{\beta}_m)$ to $H(\underline{X}_m)$ can never exceed the Sensor Channel Transmittance.

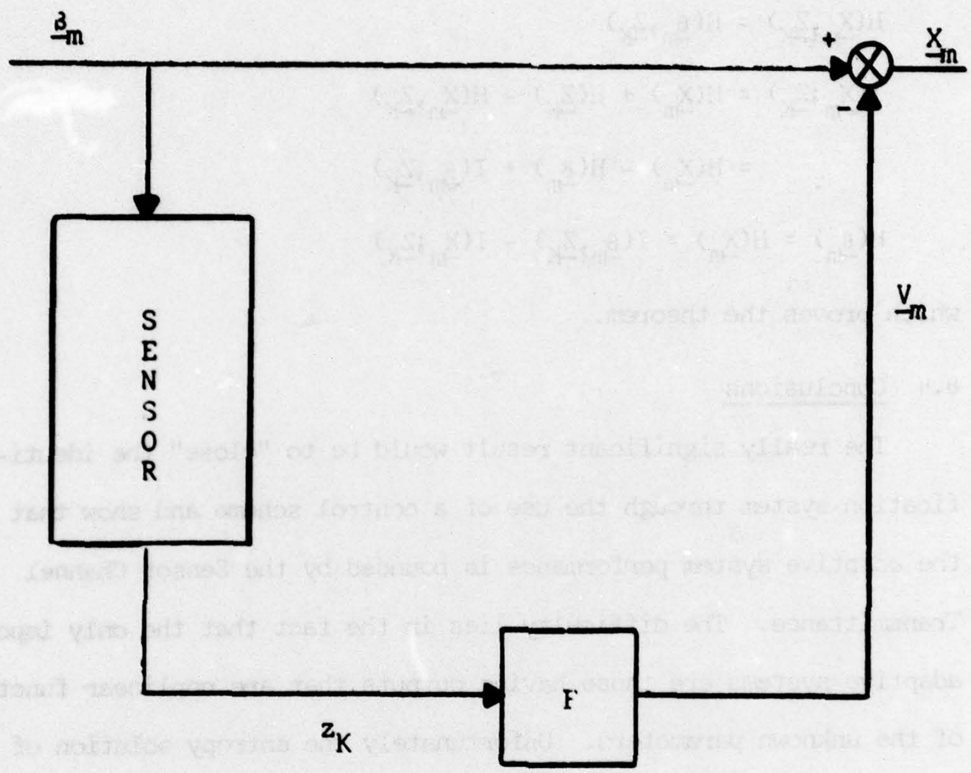


Figure 8.1. An Identification System

Proof:

$$\underline{X}_m = \underline{\beta}_m - F(\underline{Z}_K)$$

$$p_{x,z}\,(\underline{X}_m,\underline{Z}_K) = p_{\beta,z}\,(\underline{X}_m + F(\underline{Z}_K),\underline{Z}_K)$$

or

$$H(\underline{X}_m,\underline{Z}_K) = H(\underline{\beta}_m,\underline{Z}_K)$$

$$I(\underline{X}_m;\underline{Z}_K) = H(\underline{X}_m) + H(\underline{Z}_K) - H(\underline{X}_m,\underline{Z}_K)$$

$$= H(\underline{X}_m) - H(\underline{\beta}_m) + I(\underline{\beta}_m;\underline{Z}_K)$$

$$H(\underline{\beta}_m) = H(\underline{X}_m) = I(\underline{\beta}_m,\underline{Z}_K) - I(\underline{X}_m;\underline{Z}_K)$$

which proves the theorem.

8.4 Conclusions

The really significant result would be to "close" the identification system through the use of a control scheme and show that the adaptive system performance is bounded by the Sensor Channel Transmittance. The difficulty lies in the fact that the only important adaptive systems are those having outputs that are nonlinear functions of the unknown parameters. Unfortunately the entropy solution of a feedback system with that much generality is yet to be determined.

CHAPTER NINE

SUMMARY OF RESULTS AND AREAS FOR FUTURE SEARCH

9.1 Summary of Results

The application of the entropy concepts to the analysis of sampled data feed-forward, and sampled data feedback systems has led to some interesting and useful results. Through the use of the entropy as a performance measure, an information quantity - - the Sensor Channel Transmittance - - may be defined, which forms the basis of a far reaching analysis technique. For example, in the estimation problem with either real time or non-real time and with no constraints on any of the system elements or signals (other than the assumed existence of the signal entropy), it is the Sensor Channel Transmittance, or a close relative of it, which bounds the estimating system performance. Moreover, the form of the sensor need not be specified and the channel property may even be calculated experimentally if necessary. Under certain conditions, the estimation filter may be chosen in such a way that an optimum system is achieved. The performance improvement of the optimum system is then equal to the Sensor Channel Transmittance. This result is important for two reasons; first it justifies the use of entropy analysis for investigating systems, and second, entropy optimization performance limits are known independent of whether or not the necessary optimizing filter is determined in advance.

As a by product of the study leading to the major results of this dissertation, several philosophically pleasing observations were made.

The most interesting observation was derived from a study of the equation

$$H(\underline{X}) \geq H(D(\underline{Y})/\underline{Z}).$$

This expression is a direct analog to

$$\mathrm{VAR}\,\{\underline{X}\} \geq \mathrm{VAR}\,\{D(\underline{Y})/\underline{Z}\}$$

and contributes to the belief that entropy and variance are each special cases of a more general criterion function, uncertainty. The advantages of using an uncertainty function other than variance has already been demonstrated by this dissertation. For example, a monotonic decrease of error variance with time can not easily be shown, but with entropy the decrease is easily proven. At a more practical level, the form of the results are always such that they reduce to well known equations for Gaussian variables and to acceptable variance inequalities for non-Gaussian random variables. Therefore, if for no other reason, entropy analysis is justified because it always leads, very quickly, to mean square error bounds.

It is not possible to make such broad statements about feedback control systems. The most critical assumption made for the solution of this problem is that the sensor must have a linear prefilter and additive noise. Nonetheless the same conclusions regarding the importance of Sensor Channel Transmittance as a performance bound, the significance of entropy as one type of uncertainty, and the reduction in the special case to accepted Gaussian results, may still be made. The real power of the feedback entropy bounds is that not only are they independent of the feedback filter but they depend only

on the open loop behavior of the sensor. If a filter to achieve optimum performance exists, then the optimum performance achieves the entropy bound and that bound may be determined from the Sensor Channel Transmittance without first having to calculate the optimum feedback function to close the loop.

Less conclusive but just as satisfying are the results derived for adaptive control systems and continuous time systems. In both cases, results are obtained that justify the potential of entropy analysis. Theoretically, the continuous time problem has been solved and the solutions are identical to those obtained for sampled data systems, i.e., the improvement in system performance because of a feed-forward (or feedback) path is limited by the continuous Sensor Channel Transmittance of that path. The difficulty is in defining continuous time entropy and not in making use of it. Similarly, when the unknown parameters of an object are to be determined for use by a controller in an adaptive system, the ability to determine those parameters is limited by the Channel Transmittance of the sensor being used.

9.2 Areas for Future Research

There are three critical problems restricting the application of entropy analysis:

1. The assumption of a feedback sensor having a linear prefilter.
2. The assumption of the existence of a continuous time entropy function.

3. The fact that entropy measurement is not a sophisticated art.

Relaxation of the constraints on the feedback sensor are desirable because several important feedback control sensors can not be modelled exactly in the form that was used for the feedback theorem. The human operator is an example of a sensor that has no such model. It would be a great advantage if a theorem could be derived for bounding the performance of feedback systems which depends only on the channel property of the sensor. If such a generalization is not possible, it would be convenient to at least allow for nonlinear sensor prefiltering. This result in a vector context, together with Chapter Eight, would immediately solve the problem of determining the performance of a "closed loop" adaptive controller. It is expected that it is possible to make a statement of the form:

"The reduction in the output entropy of a plant which is controlled by an adaptive system is bounded by the Channel Transmittance of the sensor used to adaptively control the plant."

If the conclusions of this dissertation are accepted as being important, then there is no question that a continuous time entropy measure is required. That such a measure would be put to immediate use was demonstrated in Chapter Seven. But, until a suitable entropy function is defined, these results can not be interpreted or used effectively.

There remains only one other fundamental problem with entropy analysis and that is the newness of the concept. Systems, at present, are not usually described in terms of information quantities so a whole new set of system properties must be defined and measured. This will involve designing instruments and algorithms to aid in the determination of mutual information and entropy. This is an area of interest that is virtually uninvestigated. Certainly if entropy analysis is to achieve prominence, entropy measurements must not be neglected.

REFERENCES

1. Shannon, C. E. and W. Weaver, The mathematical theory of communication. The University of Illinois Press, Urbana, Illinois, 1963.

2. Feinstein, A., Foundations of information theory. McGraw-Hill, New York, 1958.

3. Khinchin, A. I., Mathematical foundations of information theory (translated from the Russian by R. A. Silverman and M.D. Friedman). Dover Publications, Inc., New York, 1957.

4. Wyner, A. D., "The capacity of the band-limited channel." The Bell system technical journal, 45:359-395, March, 1966.

5. Ash, R. B., "Capacity and error bounds for a time-continuous Gaussian channel." Information and control, 6:14-27, 1963.

6. Lomnicki, Z. A. and S. K. Zaremba, "The asymptotic distributions of estimators of the amount of transmitted information." Information and control, 2:260-284, 1959.

7. Abramson, N., Information theory and coding. McGraw-Hill, New York, 1963.

8. Fano, R. M., Transmission of information. The M.I.T. Press, Cambridge, Mass., 1963.

9. Ash, R. B., Information theory. Interscience Publishers, New York, 1965.

10. McMillan, B. "The basic theorems of information theory." Annals of mathematical statistics, 24:196-219, 1953.

11. Feinstein, A., "A new basic theorem of information theory." I.R.E. transactions on information theory, 4:2-22, Sept., 1954.

12. McGill, W. J., "Multivariate information transmission." I.R.E. transactions on information theory, 4:93-111, Sept., 1954.

13. Komologorov, A. N., "On the Shannon theory of information transmission in the case of continuous signals." I.R.E. transactions on information theory, 2:102-108, Dec., 1956.

14. Chover, J., "On normalized entropy and the extensions of a positive-definite functions." Journal of mathematics and mechanics, 10(6):927-945, 1961.

15. Birch, J. J., "Approximations for the entropy for functions of Markov chains." Annals of mathematical statistics, 33:930-938, Sept., 1962.

16. Gerrish, A. M., and P. M. Schultheiss, "Information rates of non-Gaussian processes." I.R.E. transactions on information theory, 10(4):265-271, Oct., 1964.

17. Golomb, S. W., "A new derivation of the entropy expressions." I.R.E. transactions on information theory, 7(3):166-167, July, 1961.

18. Karush, J., "A simple proof of an inequality of McMillan." I.R.E. transactions on information theory, 7(2):118, April, 1961.

19. Lindley, D. V., "On a measure of the information provided by an experiment." Annals of mathematical statistics, 27:986-1005, 1956.

20. DeGroot, M. H., "Uncertainty, information, and sequential experiments." The annals of mathematical statistics, 30(2): 404-419, June, 1962.

21. Kelly, J. L., "A new interpretation of information rate." The Bell system telephone journal, 35:914-926, July, 1956.

22. Elkind, J. I., "Transmission of information in simple manual control systems." I.R.E. transactions on human factors in electronics, 2(1):58-60, March, 1961.

23. Foy, Wade H., Random parameters in linear systems, Ph.D. in Engineering, the Johns Hopkins University, Baltimore, Maryland, Dec., 1961.

24. Balakrishnan, A. V., "Signal selection theory for space communications channels." Advances in communications systems (edited by A. V. Balakrishnan): 1-31. Academic Press, New York, 1965.

25. Ovseevich, I. A., and M. S. Pinsker, "Evaluation of the carrying capacity of a communication channel whose parameters are random functions of time." Radio engineering, 12(10):54-62, 1957.

26. Bishop, W. B. and B. L. Buchanan, "Message redundancy vs. feedback for reducing message uncertainty." The I.R.E. convention record, part 2:33-39, March, 1957.

27. Varshaver, B. A., "The theory of signal transmission with multiple discrete values." Radio engineering, 14(1):1-13, Nov., 1959.

28. Ovseevich, I. A. and M. S. Pinsker, "The speed of transmission of information and the carrying capacity of a multipath system and reception by the linear operator conversion method." Radio engineering, 14(3):13-29, Jan., 1960.

29. Good, I. J. and D. C. Doog., "A paradox concerning rate of information." Information and control, 1:113-126, May, 1958.

30. Swerling, P., "Paradoxes related to the rate of transmission of information." Information and control, 3:351-359, Dec., 1960.

31 Gel'Fand, I. M. and A. M. Yaglom, "Calculation of the amount of information about a random function contained in another such function." American math society translations, series 2, 12:199-246, 1959.

32. Pinsker, M. S., Information and information stability of random variables and processes. (Translated from the Russian by A. Feinstein), Holden-Day, San Francisco, 1964.

33. Hyang, Robert Y. An information theory for time continuous processes. Ph.D. in Engineering, Syracuse University, Syracuse, New York, June, 1962.

34. Balakrishnan, A. V., "Estimation and detection theory for multiple stochastic rpocesses." Journal of mathematical analysis and applications, 1:386-410, Dec., 1960.

35. Balakrishnan, A. V., "On a class of nonlinear estimation problems." I.R.E. transactions on information theory, 10(4):314-320, Oct., 1964.

36. Weiner, N., Cybernetics, 2nd edition, Wiley, New York, 1961.

37. Garner, W. R. and W. J. McGill, "The relation between information and variance analysis." Psychometrika, 21(3):219-228, Sept., 1956.

38. Swarup, Chaitanya, Some informational theoretical and empirical techniques in statistical inference. Ph.D. in Mathematics, The Univeristy of New Mexico, 1964.

39. Krasovskiy, A. A., "Entropy stability of linear continuous automatic control systems." Engineering cybernetics, 1(5): 10-16, 1963.

40. Krasovskiy, A. A., "Variation in entropy of continuous dynamic systems." Engineering cybernetics, 2(5):1-11, Sept., 1964.

41. Parzen, E., Modern probability theory and its applications, Wiley, New York, 1960.

42. Papoulis, A., Probability, random variables, and stochastic processes, McGraw-Hill, New York, 1965.

*U.S.Government Printing Office: 1979 — 657-002/668